Unsupervised Feature Learning for Heart Sounds Classification Using Autoencoder

Wei Hu¹, Jiancheng Lv¹, *, Dongbo Liu¹ and Yao Chen¹

¹College of Computer, Sichuan University, Chengdu, 610065, China;

*lvjiancheng@scu.edu.cn; Tel.: +86-138-8050-3702

Abstract. Cardiovascular disease seriously threatens the health of many people. It is usually diagnosed during cardiac auscultation, which is a fast and efficient method of cardiovascular disease diagnosis. In recent years, deep learning approach using unsupervised learning has made significant breakthroughs in many fields. However, to our knowledge, deep learning has not yet been used for heart sound classification. In this paper, we first use the average Shannon energy to extract the envelope of the heart sounds, then find the highest point of S1 to extract the cardiac cycle. We convert the time-domain signals of the cardiac cycle into spectrograms and apply principal component analysis whitening to reduce the dimensionality of the spectrogram. Finally, we apply a two-layer autoencoder to extract the features of the spectrogram. The experimental results demonstrate that the features from the autoencoder are suitable for heart sound classification.

1. Introduction

According to the World Health Organization, cardiovascular disease (CVD) is the number one cause of death globally. Cardiac auscultation [1] is the most widely used method of hearing and interpreting heart sound (HS). However, it is usually influenced by a doctor's mood, experience, or other subjective factors. A doctor needs many years of clinical experience to make an accurate diagnosis. For these reasons, it is necessary to have an intelligent, automatic HS auscultation system. Therefore, the accuracy of cardiac auscultation system should be significantly improved.

HS is complex sound caused by the vibration of the heart valve opening and closing, systolic and diastolic tendon and muscle movement, and the impact of the blood flow on the blood wall, while is low frequencies signal. Normal heart rate is 72 beats per minute, so a cardiac cycle is equal to 1/72 minutes, or 833 ms per beat [2]. A cardiac cycle can be divided into four HSs, but usually only the first two can be heard. The first HS (S1) is caused by vibration during the systole, in which the mitral and tricuspid valves shut suddenly. The second HS (S2) is mainly due to valve vibration caused by ventricular diastole, in which the aortic and pulmonary valves close suddenly.

Usually, heart murmur is the first signal of a pathological change in the heart. Heart murmurs can be seen in healthy people, but more in patients with cardiovascular disease. The location of the heart murmur in the cardiac cycle has a special pathological significance for CVD diagnosis. The heart murmur can occur in any location in an HS signal. While it is easy to locate the murmur visually from signal graph, it is extraordinary difficult to locate it in an auscultation system.

To improve the accuracy of HS classification, it is important to find an appropriate data representation (the features) of the HS signal. These features must maintain a certain similarity with the original HS, which can show characteristics of heart sound. A HS has timing, frequency, and morphology characteristics [3], which increases the difficulty of extracting features. So far, many methods for extracting features have been proposed, and achieved good results [4]. F. Yaghouby and...
A. Ayatollahi [5] present an effective arrhythmia classification algorithm using the heart rate variability (HRV) signal, basing on the Generalized Discriminant Analysis (GDA) feature reduction technique and the Multilayer Perceptron (MLP) neural network classifier. A radial wavelet neural network classifier is proposed for heart murmurs, which uses the extended Kalman filter as learning algorithm, extracting features from real cardiac cycles [6]. S-transform methods are commonly used to extract HS features. These methods are very suitable for the analysis of an unsteady signal, and their frequency features are superior to those of the Fourier transform method [3]. Four additional HS signal characteristics, namely activity, complexity, mobility, and peaks of power spectral density plots have been used as the input of a neural network [7] for classification. Higher-order cumulants are good at dealing with nonlinear signals [8] and perform HS classification very well. Moreover, [9] proposed an optimal multi-scale decomposition of a wavelet feature extraction method and used an Support Vector Machine classifier to recognize normal and abnormal HSs. Wu et al presents an approach for heart sounds identification, which ensured 95% of accuracy using wavelet transform to extract the envelope of PCG signals [10]. A new heart sound classification technique was presented in 2014, which improves in performance of the diagnostic system, by combining Linear Predictive Coding coefficients, used for future extraction, with a classifier built upon combining Support Vector Machine and Modified Cuckoo Search algorithm [11].

In recent years, deep learning (deep neural network) has achieved a huge success in image recognition, voice recognition, natural language processing, and other research fields [12]. The artificial neural network [13] is a complex information-processing model simulating the human brain's nervous system, which has the ability to extract features using a nonlinear transformation [14]. Neural networks have been successfully used in many medical applications, such as decision-making, clinical diagnosis, and prediction [15-17]. The autoencoder (AE) [18] is a kind of neural network with a three-layer structure that is based on unsupervised learning using the back propagation (BP) algorithm [19]. It makes the output value equal to the input values, so that it can extract features which is conducive to classification [20]. In this paper, we first find the envelope of the original HSs signals by calculating the average Shannon energy to extract the cardiac cycle. We convert the time-domain signals of cardiac cycles into spectrograms and reduce their dimensionality using principal component analysis [21]. Finally, we apply two-layer AEs to extract the spectrogram features. The experimental results show that the AE features are effective for HS classification.

2. Methods

Figure 1 shows the flow of our algorithm. Because a HS changes cyclically, we apply the average Shannon energy to extract an envelope for dividing the signal into cardiac cycles. We then use the cardiac cycle as data input, and use frequency spectrum analysis to obtain its spectrogram. Because of the high dimensions of the spectrogram, we adopt PCA [21] to reduce its dimensionality. Finally, features are extracted by an unsupervised learning AE neural network. After the AE processing, we use soft max to classify the HS.

![Figure 1](image-url)  
*Figure 1.* A flowchart of our algorithm. We first cut cardiac cycles from original HS signal. Then we convert each cardiac cycle into spectrogram. Spectrograms are feed to autoencoder neural network. Finally to classify.

2.1. extraction of cardiac cycle
The cardiac cycle is the basis of HS classification. The traditional method uses the HS essential elements (s1 and s2) for classification, but because a pathological HS contains murmurs, this classification was prone to error. In this paper, to avoid using the essential elements for HS classification, we use a segmentation method based on the extraction of cardiac cycles.

According to the characteristics of the HS signal, we adopt two adjacent peaks of the S1 to define a cardiac cycle, so the main task is to find the peaks. We use the average Shannon energy to extract the HS envelope, then it is simple to find the highest points. The two adjacent peaks of the envelope can be defined the cardiac cycle. Based on the Shannon energy, the normalized average Shannon energy can extract the HS envelope and is able to reduce low-amplitude noise so that the low-amplitude signal is easier to distinguish [22].

First, we use the following equation to normalize the original sound signal.

$$x_{norm}(k) = \frac{x(k)}{\max(|x(k)|)}$$  

(1)

Shannon energy emphasizes a moderate-amplitude signal, and relative to a high-amplitude signal, the low-amplitude signal is very weak. Consequently, Shannon energy is good at reducing the difference in the envelopes of the high-amplitude and low-amplitude signals. This reduction enables a low-amplitude signal to be more likely to be found.

We define the average Shannon energy as:

$$E_s = -\frac{1}{N}\sum_{i=1}^{N} x_{norm}^2(i) \log x_{norm}^2(i)$$  

(2)

Where $x_{norm}$ is the normalized signal from Equation (1) and N is the length of the signal window. In our experiment, average Shannon energy takes 0.02 s per window to calculate, and 0.01 s to step through the entire signal.

The energy envelope is calculated as follows:

$$P(t) = \frac{E_s(t) - M(E_s(t))}{S(E_s(t))}$$  

(3)

Where $M(E_s(t))$ is the average of $E_s(t)$, and $S(E_s(t))$ is the standard deviation, $P(t)$ is the normalized Shannon Energy envelope that we need.

Figure 2 shows the envelopes obtained from our experiment. Figure 2(a) shows a normal HS, including S1 and S2. In Figure 2(b), the normal HS envelope has been extracted using average Shannon energy. Figure 2(c) is a HS signal with murmurs, and Figure 2(d) is the envelope of Figure 2(c). This figure shows that the method can accurately extract the envelope of HS signal, which helps us to extract the cardiac cycle.

After obtaining the HS signal envelope, we calculate the highest point of the envelope, that is to say, we aim to obtain the highest points of S1. In our experiment, in order not to separate S1, we translate all the peaks by a distance, then extract the cardiac cycles from the envelope.
2.2. converting into Spectrogram

The original HS signals are obtained by sampling according to time, and every sample point presents the amplitude of the sound. If we directly extract features from the original signal, we can only obtain the time-domain characteristics. However, there are more HS features in the frequency-domain. Therefore, we need to find a data representation that can describe HS both in the time and frequency domains to extract the HS features accurately.

A two-dimensional map is obtained by a continuous spectral analysis of the cardiac cycle. In this map, the x-axis is time, the y-axis is frequency, and the value of each pixel reflects the energy density at the corresponding time and frequency. This map is called a sonogram or spectrogram [23]. This spectrum reflects the relationship of the HS signal over time and frequency, and is important in HS analysis.

HS is a continuous signal, and each single sample point of this signal only represents the amplitude of the HS. We put the cardiac cycle in a frame, and the frame length generally is a power of two, so a cardiac cycle is $x_n(m), n = 0, 1, \ldots, N - 1$, where m is the frame number, n represents each sampling point in a frame, and N is the length of a frame (number of sampling points in a frame). There is an overlap between frames to help extract detailed frequency information. Discrete Fourier transform (DFT) of signal $x(m)$ is:

$$X(n, k) = \sum_{n=0}^{N-1} x_n(m)e^{-j\frac{2\pi nk}{N}}$$

Where $0 \leq k \leq N - 1$, and $|X(n, k)|$ is short time amplitude spectrum estimation of $x(n)$, then we can calculate energy spectrum density function (the power spectrum function) of time m, namely $P(n,k)$:

$$P(n, k) = |X(n, k)|^2 = (X(n, k)) \otimes \text{conj}(X(n, k))$$

Where "\otimes" denotes element-wise multiplication, and "\text{conj}( )" is the function to compute conjugate value of complex number. $P(n,k)$ is two-dimensional nonnegative real function. We can obtain the spectrogram with time n as the x-axis, k as the y-axis, and the value of $P(n,k)$ as the value of the pixel.

Figure 3. Spectrograms of cardiac cycles: (a) and (b) are spectrograms of 2 cardiac cycles in the same normal HS; (c) and (d) are from HS with murmur in our experiment. The more yellow the color, the greater the level of the frequency, and conversely, the more blue color, the smaller.

Figure 3 shows that spectrograms of different cardiac cycles in a HS are very similar, so spectrum analysis does not change the HS structure. Furthermore, the spectrograms show us that a normal HS is concentrated in the low frequency, the frequency of a murmur is higher than that of a normal HS, and most of the high frequencies are zero, which conforms to HS characteristics.

2.3. extracting feature with AE

The artificial neural network is a layer system that tries to simulate the way the brain thinks, where a node represents a neuron of brain, the connections between nodes correspond to neuron connections, and the value of the connection weight reflects the degree of connection. Unsupervised learning does not need to know the data label, as it directly models the whole data set to learn useful information.
Applying unsupervised learning to HS feature extraction, we may extract features end-to-end without understanding signal processing and the characteristics of the HS signal.

AE is a typical artificial neural network structure for unsupervised learning, as shown in Figure 4. AE has three layers, the input layer, hidden layer, and output layer [18]. After training convergence, we obtain a representation of the input x, namely, the value of the hidden layer. The representation is not controlled manually, but learned from all training data via a neural network, where \( w^{(1)} \) is the feature.

![Figure 4. A autoencoder structure: x is input, y is output, and w^{(1)} is weight of the first layer, w^{(2)} is weight of the second layer. Each small circle represents a neuron, and the straight line represents the connection between the neurons. Neurons on the same layer are not connected.](image)

The label of HS is its target output. So we use the BP algorithm to adjust its parameters and obtain the weight of each layer. To do so, we obtain a representation of input x (the value of the hidden layer), and \( w^{(1)} \) is defined as the feature. AE is a kind of neural network that tries to reconstruct the original input signal, which is the same as the restricted Boltzmann machine [25]. That is to say, AE tries to learn the function \( h_{w,b}(x) \approx x \).

Through forward propagation, we calculate the value of the hidden layer and output:

\[
z^{(2)} = x \ast W^{(1)} + b^{(1)}
\]

\[
a^{(2)} = f(a^{(2)})
\]

\[
z^{(3)} = a^{(2)} \ast W^{(2)} + b^{(2)}
\]

\[
y = f(z^{(3)})
\]

Where, \( a^{(2)} \) is the value of hidden layer, \( a^{(1)} = x \), \( a^{(3)} = y \); \( b^{(l)} \) denotes the bias of \( l \) layer; \( z^{(l)} \) is the input of \( l \) layer; \( f(x) \) denotes the activation function; \( y \) is the output of network. Bias is also the training parameter, which has the effect of regulating the network and enhancing the adaptability of the network. We use sigmoid function as our activation function:

\[
f(x) = \frac{1}{1 + \exp(-x)}
\]

We compute all the activations of the nodes using the equations above that describe the forward propagation step. As stated above, AE tries to learn the function \( h_{w,b}(x) \approx x \), that is to say, AE tries to make \( y \) equal to \( x \), or reduce the reconstruction error. We define the cost function:

\[
J(x,y) = \frac{1}{2} \| y - x \|^2
\]

Combining equations (8) and (9):

\[
J(W,b; x) = \frac{1}{2} \| f(f(x \ast W^{(l)} + b^{(l)}) \ast W^{(2)} + b^{(2)}) - x \|^2
\]

In order to make \( J \) be smallest, we apply gradient descent to update the network parameters \( W \) and \( b \):

\[
W^{(l)}_{ij} = W^{(l)}_{ij} - \alpha \frac{\partial}{\partial W^{(l)}_{ij}} J(W,b; x)
\]

\[
b^{(l)}_i = b^{(l)}_i - \alpha \frac{\partial}{\partial b^{(l)}_i} J(W,b; x)
\]
Where \( \alpha \) is learning rate, which is usually a very small value. Because \( \frac{\partial}{\partial W_{ij}} J(W, b; x) \) and \( \frac{\partial}{\partial b_i} J(W, b; x) \) are very difficult to compute directly, we apply back propagation (BP) \[26\] to calculate them. Using the above equations and algorithm, the network can update \( W \) and \( b \).

We impose some constraints on AE to help the network learn a better representation, such as the number of nodes in a hidden layer. In addition, we introduce the sparse constraint \[27\], which is used frequently in AE. Assume that when output of the node is close to 1, the node is thought of as activated, and when the node output is close to 0, we consider the node to be inactivated. The sparse constraint aims to inactivate most of hidden nodes.

Recall \( a_j^{(2)} \) denotes the activation of hidden node \( j \), then we compute the average activation of the hidden.

\[
\hat{\rho}_j = \frac{1}{m} \sum_{i=1}^{m} a_j^{(2)}
\]  

(15)

Obviously, we would like to make \( \hat{\rho}_j \) close to 0, namely \( \hat{\rho}_j = \rho \), where \( \rho \) is a small value closing to 0 (\( \rho = 0.05 \)). To achieve the constraint, we add an penalty factor to \( J \) as follow:

\[
KL(\rho || \hat{\rho}_j) = \sum_{j=1}^{m} \left( \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{\rho}_j} \right)
\]  

(16)

Where \( m \) is the number of hidden nodes. Now our cost function is:

\[
J_{sparse}(W, b) = J(W, b) + \beta \sum_{j=1}^{m} KL(\rho || \hat{\rho}_j)
\]  

(17)

Multi-layer artificial neural networks have excellent ability to learn feature extraction, and the features learned have more essential description to the data, which facilitates visualization and classification \[28\]. The paper also proposes that the training difficulty of multi-layer neural network can be overcome by using layer-wise pre-training. Since AE is a simple network of three layers, we increase the number of layers to extract more descriptive features in our works.

As shown in Figure 5, we first trained an AE network, and after the network convergence, we used the AE hidden layer as the input layer of another AE. Training step-by-step reduces the training difficulty of a multi-layer network.

Fine-tuning \[18\] is a common tip in AE training. It can improve the accuracy of classification, and has been widely used in neural network. It can improve the classification accuracy and has been widely used in neural networks. It is likely that we can find a good initial point via fine-tuning, then globally train the neural network.

3. Results

3.1. Experiment preparation
Accuracy is a simple and effective experimental evaluation metric, which is the ratio of the number of classification correct and the number of all data.

As far as we know, there is no publically available standard HS dataset so far. In our experiment, most HS signals were collected by ourselves through Bluetooth stethoscopes. In addition, data from the Internet is also included. We integrated these data and divided them into two categories, namely, normal HS and abnormal HS. This is a simple and effective classification, and using this classification, we can quickly diagnose whether the HS is normal or abnormal. In our dataset, there are positive class 1126 and negative class 1525. Of these, thirty percent serve for test sets, seventy percent for training sets.

Furthermore, according to the location of the murmur in a cardiac cycle, we mark an HS signal into seven classes (normal, early systolic murmur, mid-systolic murmur, late systolic murmur, whole systolic murmur, diastolic murmur, and systolic click). The classification method pinpoints the location of murmur, which is more useful for CVD diagnosis. Of course, it greatly increased the difficulty of classification. Similarly, thirty percent serve for test sets, seventy percent for training sets.

### 3.2. Experiment results

The network structure we adopt is described above. In the first experiment, The first AE (1-AE) had 196 nodes for the input layer and 100 nodes for the hidden layer nodes. After AE training convergence, we gain the feature (the hidden layer). Then we used softmax for classification, which has fine-tuning structure. Table 1 shows that the accuracy has reached 99.13%.

Further, we then design the second AE with 100 nodes for the input layer and 64 nodes for the hidden layer nodes. Through layer-wise pre-training, we obtained a better representation of the data. In the experiment, the accuracy reached 99.48%, which is superior to only one-layer AE. The experiment results show that the feature of HSs extracted by multilayer network has more essential expression to HS, and is more conducive to classification.

| method | accuracy in train set(%) | accuracy in test set(%) |
|-------|--------------------------|------------------------|
| 1-AE  | 99.5                     | 99.13                  |
| 2-AE  | 100                      | **99.48**              |
| RBF   | 68.6                     | 72.63                  |
| MLP   | 99.93                    | 98.83                  |
| SVM   | 100                      | 97.83                  |

Table 2 also presents some representative experimental results of other methods. This table shows that the radial basis function (RBF) neural network performs very poorly, indicating that the RBF has difficulty learning an effective representation from the lower-dimensional data reduced using PCA. The SVM performs very well on the training set, but the test set results are poorer than those of the neural network method. A support vector machine constructs a hyperplane, or in a high or infinite dimensional space, which can be used for classification and regression. In addition, the performance of MLP is good (98.83%), but the performance of AE is better still. It is proved that the accuracy of Heart Sound classification can be improved by using the feature extracted by AE in the same network structure.

In our second experiment, we recognized the seven-class data. Apparently, The seven classification is more difficult than the two classification. If our method performs well on the seven category, it is
sufficient to show that our approach is successful. Again, we adopted a two-layer AE for feature extraction and classification.

Table 2 shows that the accuracy of seven classification is indeed lower than that of the two classification several. And we can see that the feature extraction of AE based on unsupervised learning for HS classification is better than that of other methods (the accuracy is the highest). Above all, the high accuracy (98.55%) proves that we can find the exact location of the heart murmur by our method.

4. Discussion
Traditional methods first extract HS features, then process and classify them. A neural network provides an end-to-end model that combines feature extraction and classification. Each layer of the model represents a kind of feature. Further, if the layers are deeper, the features are more abstract. We can interpret each layer of the neural network as a data space, and transform the data into a different space. The aim of the neural network is to convert data into a space suitable for classification. Our experimental results also show that the neural network classification of HS has achieved excellent results. Enhancing the accuracy of HS classification is a precondition for establishing an automatic auscultation system. Using the increasingly popular deep neural network method, the accuracy is already close to 100%. Therefore, it is possible that we can accurately locate the location of the heart murmur, which can help the patient find his heart problems at an early date.

To obtain a good classification model, deep neural network requires a large amount of data for training. Now that big data is becoming more available, data are not the problem. A portable automatic auscultation system with high accuracy could soon appear in everyone's home. At that point, patients will be able to listen to their own HS signals anytime and anywhere, and as early as possible be able to protect their own hearts.

5. Conclusions
In this paper, we proposed a set of HS classification methods based on autoencoder. First, cardiac cycles are extracted by extracting the HS envelope to avoid the need to identify the HS composition. A spectrogram that consists of the time-domain and frequency-domain characteristics of cardiac cycles is fed into an AE. We use an unsupervised learning multi-autoencoder to extract high-level features. In the first experiment, the proposed method recognized whether the HS was normal or abnormal quickly and accurately. In the second experiment, the proposed method was able to recognize the location of the murmur, which is important for CVD diagnosis. These experimental results show that the proposed method performs better than other methods in accuracy. So, the intelligent and automatic HS auscultation system is realized in the future, and it could help doctors or patients better identify heart problems.

Acknowledgments
This work was supported by the National Science Foundation of China (Grant No. 61375065 and 61625204), partially supported by the State Key Program of National Science Foundation of China (Grant No. 61432012 and 61432014).

References
[1] BENDER J R. Yale university school of medicine heart book. New Yourk:William Morrow and Company 1992
[2] I. Khurana. Textbook of Medical Physiology. Elsevier, Noida, UP, 2006, pp. 1204-1209.
[3] G.Amit, N.Gavriely, J.Lessicick, N. Intrator. Acoustic Indices of Cardiac Functionality. International Conference on Bio-inspired System and Signal Processing (BIOSINGALS) 2008, vol 2, pp 77-83
[4] P. Sengottuvelan, R. Arulmurugan. Object classification using substance based neural network. Math. Prob. Eng 2014
[5] F. Yaghouby, A. Ayatollahi, R. Soleimani. Classification of Cardiac Abnormalities Using Reduced Features of Heart Rate Variability Signal. World Applied Sciences Journal 6 (11), 2009, pp. 1547-1554
[6] J.E. Guillermo, L.J.R. Castellanos, E.N. Sanchez, A.Y. Alanis. Detection of heart murmurs based on radial wavelet neural network with kalman learning. Neurocomputing 2015, 164, pp. 307-317

[7] Ibrahim Abdel-Motaleb, Rohit Akula. Artificial Intelligence Algorithm for Heart Disease Diagnosis using Phonocardiogram Signals. IEEE transaction on biomedical engineering 2012, Vol. 51, NO. 7, pp. 1196-1202

[8] C. Ahlstrom, P. Hult, P. Rask, J.E. Karlsson et al. Feature extraction for systolic heart murmur classification. Annals of Biomedical Engineering 2006, 34 (November (11)), pp. 1666-1677

[9] Wang Y, Li W.Z, Zhou J.L, Li X.H, Pu Y.F. Identification of the normal and abnormal heart sounds using wavelet-time entropy features based on OMS-WPD. Futur. Gener. Comput. Syst 2014, 37, 488-495

[10] Wu JB, Zhou S, Wu Z, Wu XM. Research on the method of characteristic extraction and classification of phonocardiogram. In: Systems and Informatics (ICSAI), 2012 International Conference on, pp. 1732–1735

[11] G. Redlarski, D. Gradolewski, A. Palkowski, "A system for heart sounds classification", PloS one 2014, pp. e112673

[12] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015, 521, 436-444

[13] J. C. Lv, K. K. Tan, Z. Yi, S. Huang. A family of fuzzy learning algorithms for robust principal component analysis neural networks. IEEE Trans. Fuzzy Syst. 2010, vol. 18, no. 1, pp. 217-226

[14] X. Yang, J. Zheng. Artificial neural networks. Handbook of Research on Geoinformatics 2009, p. 122

[15] Alkan. A, Koklukaya. E, Subasi. A. Automatic seizure detection in EGG using logistic regression and artificial neural network. Neurosci

[16] Sawa. T, Ohno-Machado. L. A neural network-based similarity index for clustering DNA microarray data. Comput. Biol. Med 2003, 3, 1-15

[17] Tan. A-H, Pan. H. Predictive neural network for gene expression data analysis. Neural Networks 2005, 18, 297-306

[18] P. Vincent, H. Larochelle, Y. Bengio, P.-A. Manzagol. Extracting and composing robust features with denoising autoencoders. Proceedings of the Twenty-fifth International Conference on Machine Learning (ICML'08) 2008, (W. W. Cohen, A. McCallum, and S. T. Roweis, eds.), pp. 1096-1103, ACM

[19] LeCun. Y, Boser. B, Denker. J.S et al. Handwritten digit recognition with a back-propagation network. Advances in Neural Information Processing Systems 1990, 2, 396-404, Morgan Kaufman

[20] A. Ng. Sparse autoencoder. CS294A Lecture notes 2011, Stanford University, pp. 72.

[21] Jian Cheng Lv, Zhang Yi, Y. Li. Non-divergence of stochastic discrete time algorithms for PCA Neural Network. IEEE Transactions on Neural Networks and Learning Systems 2015, vol. 26, no. 2, pp. 394-399

[22] H Liang, S Lukkarinen, I. Hartimo. Heart Sound Segmentation Algorithm Based on Heart Sound Envelogram. Computers in Cardiology 1997 24: 104

[23] Greenberg. S., Kingsbury. B. The modulation spectrogram: In pursuit of an invariant representation of speech. Proc. ICASSP 1997, pp.1647-1650

[24] B. Fritzke. Growing cell structures -a selforganizing network for unsupervised and supervised learning. Neural Networks 1994, 7(9):1441-1460

[25] G. E. Hinton. A practical guide to training restricted Boltzmann machines. Tech. Rep. 2010. UTML TR 2010-003

[26] Rumelhart D E and McClelland J L. Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Cambridge, MA: MIT 1986,1

[27] A. Ng. Sparse autoencoder. CS294A Lecture notes 2011, Stanford University, pp. 72

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