Artificial Ingelligence in Medicine, 2002, vol.24, no.1, pp.25-36. @Elsevier

Lung Cancer Cell Identification Based on Artificial Neural Network Ensembles

Zhi-Hua Zhou*, Yuan Jiang, Yu-Bin Yang, Shi-Fu Chen

National Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, P.R.China

Abstract

An artificial neural network ensemble is a learning paradigm where several artificial neural networks are jointly used to solve a problem. In this paper, an automatic pathological diagnosis procedure named Neural Ensemble based Detection (NED) is proposed, which utilizes an artificial neural network ensemble to identify lung cancer cells in the images of the specimens of needle biopsies obtained from the bodies of the subjects to be diagnosed. The ensemble is built on a two-level ensemble architecture. The first-level ensemble is used to judge whether a cell is normal with high confidence where each individual network has only two outputs respectively *normal cell* or *cancer cell*. The predictions of those individual networks are combined by a novel method presented in this paper, i.e. *full voting* which judges a cell to be normal only when all the individual networks judge it is normal. The second-level ensemble is used to deal with the cells that are judged as cancer cells by the first-level ensemble, where each individual network has five outputs respectively *adenocarcinoma*, *squamous cell carcinoma*, *small cell carcinoma*, *large cell carcinoma*, and *normal*, among which the former four are different types of lung cancer cells. The predictions of those individual networks are combined by a prevailing method, i.e. *plurality voting*. Through adopting those techniques, NED achieves not only a high rate of overall identification but also a low rate of false negative identification, i.e. a low rate of judging cancer cells to be normal ones, which is important in saving lives due to reducing missing diagnoses of cancer patients.

Keywords: Artificial neural networks; Pattern recognition; Image processing; Computer-aided medical diagnosis; Expert system

1. Introduction

Lung cancer is one of the most common and deadly diseases in the world. Detection of lung cancer in its early stage is the key of its cure. In general, measures for early stage lung cancer diagnosis mainly includes those utilizing X-ray chest films, CT, MRI, isotope, bronchoscopy, *etc.*, among which a very important measure is the so-called pathological diagnosis that analyzes the specimens of needle biopsies obtained from the bodies of the subjects to be diagnosed. At present, the specimens of needle biopsies are usually analyzed by experienced pathologists. Since senior pathologists are rare, reliable pathological diagnosis is not always available.

During the last decades, along with the rapid developments of image processing and pattern recognition techniques, computer-aided lung cancer diagnosis attracts more and more attention. Many achievements have already been attained. Examples are as follows. Chiou *et al.* [5] designed an artificial neural network based hybrid lung cancer detection system named HLND, which was used to improve the accuracy of diagnosis and

* Corresponding author. Tel.: +86-25-359-3163; fax: +86-25-330-0710.
E-mail addresses: zhouzh@nju.edu.cn (Z.-H. Zhou), jy@ai.nju.edu.cn (Y. Jiang), yyb@ai.nju.edu.cn (Y.-B. Yang), chensf@nju.edu.cn (S.-F. Chen).

the speed of lung cancerous pulmonary radiology. Lin *et al.* [16] developed a system based on a parameterized two-level convolution artificial neural network and on a special multi-label output encoding procedure, which was used in the diagnosis of lung cancer nodules found on digitized chest radiographs. Hayashibe *et al.* [11] proposed an automatic method based on the subtraction between two serial mass chest radiographs, which was used in the detection of new lung nodules. Kanazawa *et al.* [14] presented a system that extracted and analyzed features of the lung and pulmonary blood vessel regions and then utilized defined rules to perform diagnosis, which was used in the detection of tumor candidates from helical CT images. Mori *et al.* [20] proposed a procedure to extract bronchus area from 3-D chest X-ray CT images, which was used in a virtualized bronchoscopy system. Penedo *et al.* [22] developed a system that employed an artificial neural network to detect suspicious regions in a low-resolution image and employed another artificial neural network to deal with the curvature peaks of the suspicious regions, which was used in the detection of lung nodules found on digitized chest radiographs.

In this paper, based on the recognition of the power of artificial neural network ensemble, an automatic pathological diagnosis procedure named Neural Ensemble based Detection (NED) is proposed and realized in an early stage Lung Cancer Diagnosis System (LCDS). NED utilizes an artificial neural network ensemble to identify cancer cells in the images of the specimens of needle biopsies obtained from the bodies of the subjects to be diagnosed. The ensemble used is built on a specific two-level ensemble architecture and a novel prediction-combining method, which achieves not only a high rate of overall identification but also a low rate of false negative identification, i.e. a low rate of judging cancer cells to be normal ones.

The rest of this paper is organized as follows. In Section 2, artificial neural network ensemble technology is briefly introduced. In Section 3, the whole LCDS system in which NED is realized is shortly described. In Section 4, the NED procedure is proposed and some experimental comparisons are presented. Finally in Section 5, conclusions are drawn and several issues for future works are indicated.

2. Artificial neural network ensembles

From the brief review on computer-aided lung cancer diagnosis presented in Section 1, it is obvious that artificial neural networks have already been widely exploited in this area. Generally speaking, artificial neural networks are very useful in pattern recognition. Hornik *et al.* [12] showed that feedforward artificial neural networks with one hidden layer can approximate any functions in any accuracy. However, until now there is no rigorous theory indicating how to do such error-free approximation. Therefore whether an artificial neural network based application will be successful or not is almost fully determined at present by that who is the user. In general, the more experiences the user has on artificial neural networks, the more chances the application will have in gaining success. Unfortunately, in real-world applications the users are often those with little knowledge on neural computing. Therefore the fruits that artificial neural network techniques may obtain do not always appear.

In the beginning of the 1990s, Hansen and Salamon [9] showed that the generalization ability of an artificial neural network system can be significantly improved through ensembling artificial neural networks, i.e. training several artificial neural networks and combining their predictions. Later, Hansen *et al.* [10] applied artificial neural network ensemble to handwritten digit recognition and attained astonishing good results whose accuracy is 20-25% better than that of the best individual artificial neural network. Subsequently there appears a hot wave in investigating artificial neural network ensembles, which lasts up to present. Much work has been put in analyzing why and how artificial neural network ensembles work. The classical one is Krogh and Vedelsby ’s work [15], in which they derived that the generalization ability of the applied ensemble is determined by the average generalization ability and the average ambiguity of the individual artificial neural networks that constitute the ensemble. As to the definition of an artificial neural network ensemble, the most accepted one is that an artificial neural network ensemble is a collection of a (finite) number of artificial neural networks that are

trained for the same task [27]. However, some researchers prefer another definition, i.e. an artificial neural network ensemble consists of a set of individually trained artificial neural networks whose predictions are combined when classifying novel instances [21]. Work satisfying the latter definition can be traced up to what was done in the NESTOR system by Cooper [6].

An ensemble approach can be viewed as comprising two kinds of methods, i.e. a method for generating individual artificial neural networks and a method for combining individual predictions. Schapire's *Boosting* [23] and Breiman's *Bagging* [1] are prevailing methods for generating individual networks. *Simple averaging* and *weighted averaging* are prevailing methods for combining individual predictions of regression estimators. *Majority voting* and *plurality voting* are prevailing methods for combining individual predictions of classifiers. *Majority voting* judges a prediction to be the final output if more than half of the individual networks vote to the prediction. *Plurality voting* judges a prediction to be the final output if the prediction ranks first according to the number of votes.

Much work has been done in designing ensemble approaches. Examples are as follows. Maclin and Shavlik [18] utilized competitive learning to generate individual networks and then combined their outputs via simple averaging. Taniguchi and Tresp [28] designed *variance-based weighting* and *variance-based Bagging*, and experimentally found that the improvement in performance that could be achieved by averaging depended critically on the degree of regularization which was used in training the individual networks. Sharkey *et al.* [25] ensembled both an ensemble biased to false negative and an ensemble biased to false positive to improve the accuracy while lowering the false positive rate. Liu and Yao [17] proposed negative correlation learning to encourage specialization and cooperation among the individual networks, where all the individual networks were trained simultaneously through the correlation penalty terms in their error functions. Chan [3] presented weighted least square ensemble that did not require that the individual networks were independent. More developments in this area can be found in [24].

Since artificial neural network ensembles work remarkably well and are easy to be used, they are regarded as a promising methodology that can profit not only experts in artificial neural network research but also engineers in real-world applications. Besides Hansen *et al.* 's work in handwritten digit recognition [10], artificial neural network ensembles have already been applied to many real-world domains such as scientific image analysis [4], face recognition [8, 13], OCR [19], seismic signals classification [26], breast cancer diagnosis [25], and in-vitro fertilization treatment [7].

3. Lung cancer diagnosis system

X-ray chest films are valuable in lung cancer diagnosis. However, there are cases where subsequent examination should be performed to increase the reliability of diagnoses. For example, if a patient has suffered pneumonia before, some tumors may appear in abnormal lung areas so that they are difficult to be distinguished from cicatrices. In this paper, an early stage lung cancer diagnosis system named LCDS, in which NED is realized, is used to accomplish this task through utilizing needle biopsy. In other words, LCDS is used to deal with cases that could not be clearly diagnosed by X-ray chest films. Since LCDS follows the analysis of X-ray chest films, the subjects diagnosed by it are much probability of being lung cancer patients. In fact the proportion of positives, i.e. lung cancer patients, is about 70-80% among all the subjects diagnosed by LCDS.

LCDS system is depicted in Figure 1. The hardwares mainly include a light microscope, a digital camera, and an image capturer. The digital camera is mounted on the light microscope whose power of amplification is 400, which picks up the video signals of the haematoxylin-eosine (*HE*) stained specimens of needle biopsies obtained from the bodies of the subjects to be diagnosed. Those signals are captured by the image capturer and then transformed to 24 bit RGB color images that are used in cancer cell identification.

LCDS system uses convolution filters with *Gaussian* pulse to smooth the cell images. The contrast and color of the images are enhanced. Then the nucleuses in the images are segmented by thresholding. All of those are

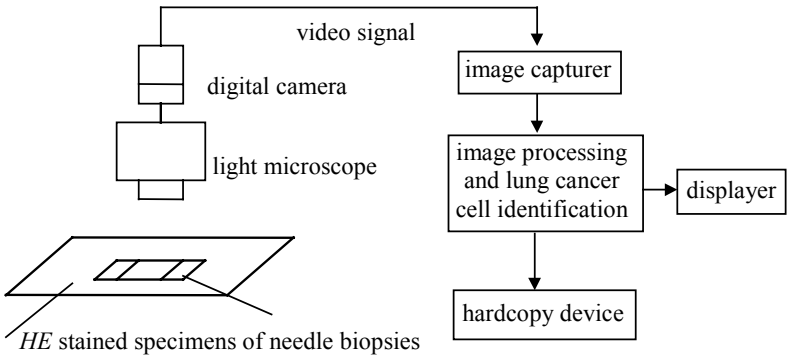

Figure 1. Lung cancer early stage diagnosis system LCDS

simple digital image processing techniques [2]. After that, LCDS utilizes morphologic and colorimetric techniques to extract features from the images of the nucleuses. The extracted morphologic features include the perimeter, area, roundness, and rectangleness of the nucleus. The extracted colorimetric features include the red component, green component, blue component, illumination, saturation, difference between red and blue components, and proportion of blue component of the nucleus. Also the red component, green component, and blue component of the entire image are included as colorimetric features. On this basis, a lung cancer cell identification module is employed to analyze those features to judge whether cancer cells exist in the specimens or not. Moreover, if there are cancer cells, the cancer cell type is identified. The entire diagnosis process of LCDS is shown in Figure 2.

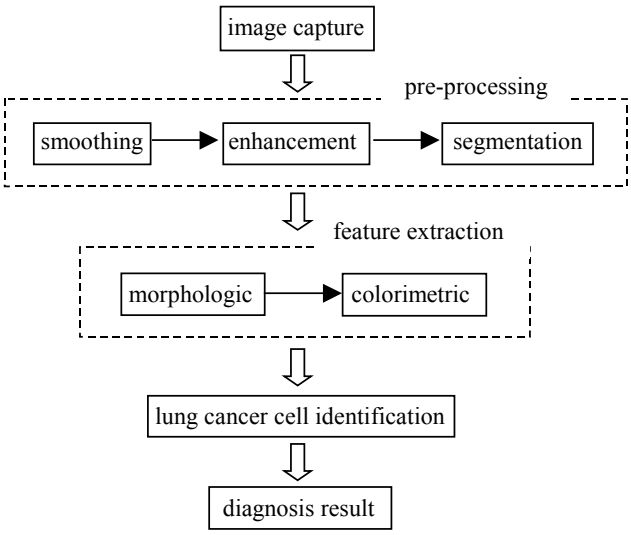

Figure 2. Diagnosis process of LCDS

4. Lung cancer cell identification

Lung cancer is a deadly disease and if a patient is correctly diagnosed in early stage there is a big chance for him or her to be cured. In other words, diagnosing a lung cancer patient to be a healthy person in the early stage is a serious mistake that results in the loss of a life. Therefore in computer-aided lung cancer diagnosis systems the rate of false negative identification, i.e. the rate of judging patients to be healthy persons, should be kept as low as possible along with the endeavor to improve the rate of overall identification as high as possible.

The data set we used comprises 552 cell images obtained from specimens of needle biopsies, whose labels

were confirmed by open biopsy. About 75% of those 552 images belong to cancer cells. The distribution is based on the fact that the proportion of lung cancer patients is about 70-80% among all the subjects diagnosed by LCDS. Among those cancer cells, about 32% are *adenocarcinoma*, about 38% are *squamous cell carcinoma*, about 22% are *small cell carcinoma*, and about 8% are *large cell carcinoma*. All the distributions approximate real-world proportion of those lung cancer types. We run 5-fold cross validation on the data set. In detail, we divide the data set into five subsets with similar size, where the proportion of different classes in each subset is similar to that in the original data set. Then we run each experiment for five times, each time using the union of four subsets as training set to train the lung cancer cell identification module depicted in Figure 2, and using the remaining subset as test set to test the trained module to see how well it works.

4.1. Results of single artificial neural networks

At first we realize the lung cancer cell identification module with a single artificial neural network. The neural algorithm we used is FANNC [29], which is a fast adaptive neural classifier that performs one-pass incremental learning with fast speed and high accuracy and does not require the user manually set up the number of hidden units.

To efficiently utilize the training data, in each experiment we employ *Bagging* [1] to generate five training sets from the original training set. The size of those five training sets is equal to that of the original training set. Then we use those training sets to train five FANNC networks each has five output units representing *adenocarcinoma*, *squamous cell carcinoma*, *small cell carcinoma*, *large cell carcinoma*, and *normal*. The experimental results are tabulated in Table 1, where each column records the average test result of the five trained networks in a experiment except that the column titled *Ave.* records the average value of those five experiments, i.e. the results of 5-fold cross validation. There are three error measures. *Err* measures the rate of overall false identification that is computed through dividing the number of false identified images by the number of test images. Err_{fn} measures the rate of false negative identification that is computed through dividing the number of images that are cancer cells but are erroneously identified as normal cells by the number of test images. Err_{fp} measures the rate of false positive identification that is computed through dividing the number of images that are normal cells but are erroneously identified as cancer cells by the number of test images. Note that there are cases where cancer cells are labeled with wrong cancer types, e.g. a cell belongs to *adenocarcinoma* is erroneously identified as *squamous cell carcinoma*, therefore the sum of Err_{fn} and Err_{fp} does not equal to *Err*.

Table 1
Experimental results of the single artificial neural networks

	Exp1	Exp2	Exp3	Exp4	Exp5	Ave.
Err	48.2%	44.5%	45.9%	41.4%	47.3%	45.5%
Err_{fn}	19.1%	14.5%	20.7%	15.3%	17.3%	17.4%
Err_{fp}	20.9%	17.3%	18.9%	16.2%	21.8%	19.0%

It is obvious that the lung cancer cell identification module realized by single artificial neural networks is highly unsatisfactory because the overall identification accuracy is even worse than 60%. Moreover, when we look into the trained networks, we find that in most experiments (*Exp1*, *Exp2*, *Exp3*, and *Exp5*) the network with the best overall false identification rate is different to that with the best false negative identification rate. This illuminates that for single artificial neural networks the endeavor to get the best accuracy of overall identification is not bound to obtain the best accuracy of false negative identification.

4.2. Results of two kinds of ensemble

Secondly, by recognizing the power of artificial neural network ensembles, we try to realize the lung cancer

cell identification module with two kinds of ensemble approaches.

In the first kind of ensemble, in each experiment we train five FANNC networks and then combine their predictions through *plurality voting* that is introduced in Section 2. The training sets of those FANNC networks are also generated via *Bagging* as described in Section 4.1. If more than one individual prediction rank first according to the number of votes, e.g. two individual networks predict *normal* while two individual networks predict *adenocarcinoma*, the identification is labeled as wrong. Moreover, if the disputed cell is a cancer cell, then both the number of false identified images and the number of false negatively identified images increase by one; if the disputed cell is a normal cell, then both the number of false identified images and the number of false positively identified images increase by one. The experimental results are tabulated in Table 2, where each column records the test result of the ensemble in a experiment except that the column titled *Ave.* records the average value of those five experiments, i.e. the results of 5-fold cross validation.

Table 2
Experimental results of the first kind of artificial neural network ensemble

	Exp1	Exp2	Exp3	Exp4	Exp5	Ave.
Err	24.5%	19.1%	21.6%	16.2%	22.7%	20.8%
Err_{fn}	8.2%	5.5%	9.0%	6.3%	7.3%	7.3%
Err_{fp}	9.1%	6.4%	7.3%	5.4%	10.0%	7.6%

Comparing Table 2 with Table 1, it is obvious that all the three error measures are significantly improved through utilizing ensemble instead of single artificial neural network to realize the lung cancer identification module. However, the rate of false negative identification is about 7.3%, which is not very satisfactory with respect to our objective.

The second kind of ensemble approach is a variation of that proposed by Sharkey *et al.* [25]. The objective of Sharkey *et al.* was to lower the rate of false positive identification. In their approach, two artificial neural network ensembles were trained, among which one ensemble was trained biased to benign diagnoses by letting negative examples dominating the training sets while the other ensemble was trained biased to malign diagnoses by letting positive examples dominating the training sets. Since the original approach was designed for problems with two output classes, we scale it by training five ensembles each comprises five FANNC networks that are trained biased to one output class by letting the examples of the biased class occupying 75% proportion in the training sets. Then the outputs of the five ensembles are combined in a way similar to winner-take-all competition. The experimental results are tabulated in Table 3, where each column records the test result of the scaled approach in a experiment except that the column titled *Ave.* records the average value of those five experiments, i.e. the results of 5-fold cross validation.

Table 3
Experimental results of the second kind of artificial neural network ensemble

	Exp1	Exp2	Exp3	Exp4	Exp5	Ave.
Err	17.3%	12.7%	13.5%	9.0%	15.5%	13.6%
Err_{fn}	7.3%	6.4%	8.1%	5.4%	6.4%	6.7%
Err_{fp}	3.6%	1.8%	2.7%	1.8%	4.5%	2.9%

Comparing Table 3 with Table 2, it is obvious that *Err* is significantly improved through utilizing the second kind of ensemble instead of the first kind of ensemble to realize the lung cancer identification module. The main reason is that the second kind of ensemble significantly lowered Err_{fp}. However, Err_{fn} has only been slightly improved.

4.3. Neural ensemble based detection

In order to significantly lower the false negative identification rate while attaining high overall identification rate, NED is proposed. NED employs a specific two-level ensemble architecture. The first-level ensemble is utilized to judge whether a cell is a cancer cell with high confidence. For this purpose, we devise a novel prediction combining method, i.e. *full voting*. Different to current prevailing methods such as *majority voting* and *plurality voting*, *full voting* holds a very strong claim that a prediction is judged as the final output only when all the individual networks hold the prediction. This is analogical to the situation that several pathologists are diagnosing a subject. The subject is judged to be healthy only when all the pathologists agree that he or she is healthy. Since the claim is very strong, we believe that *full voting* can only be used in tasks where only two output classes exist, among which one output is more important than the other.

To use *full voting*, we should modify the training data so that the number of output class is reduced from five to two. This is done by merging the cancer classes *adenocarcinoma*, *squamous cell carcinoma*, *small cell carcinoma*, and *large cell carcinomal* to a big class, i.e. *cancer cell*. Therefore 75% examples in the training sets belong to the class *cancer cell* and the rest 25% belong to the class *normal cell*. Then *Bagging* is employed again to train five individual FANNC networks each has two output units.

The cells that are judged to be cancer cells by the first-level ensemble are passed to the second-level ensemble that is responsible to report the type of the cells. Here we also employ *Bagging* to generate five FANNC networks each has five output units and then using *plurality voting* to combine the individual predictions. If more than one individual prediction rank first according to the number of votes, the identification is labeled as wrong. Moreover, if the disputed cell is a cancer cell, then both the number of false identified images and the number of false negatively identified images increase by one; if the disputed cell is a normal cell, then both the number of false identified images and the number of false positively identified images increase by one. The experimental results of NED are tabulated in Table 4, where each column records the test result of NED in a experiment except that the column titled *Ave.* records the average value of those five experiments, i.e. the results of 5-fold cross validation.

Table 4
Experimental results of NED

	Exp1	Exp2	Exp3	Exp4	Exp5	Ave.
Err	15.5%	10.0%	11.7%	8.1%	12.7%	11.6%
Err_{fn}	3.6%	1.8%	3.6%	1.8%	2.7%	2.7%
Err_{fp}	5.5%	3.6%	4.5%	2.7%	6.4%	4.5%

Comparing Table 4 with Table 1 and Table 2, it is obvious that all the three error measures are significantly improved through utilizing NED instead of single artificial neural networks or the first kind of ensemble to realize the lung cancer identification module. Comparing Table 4 with Table 3, it is obvious that although the second kind of ensemble approach is better on Err_{fp}, NED is significantly better on Err_{fn}. Moreover, NED is better on *Err*. Therefore we believe that NED is the best approach in realizing the lung cancer cell identification module of LCDS.

In summary, the flowchart of NED is depicted in Figure 3. A point worth noting is that in NED although both the first-level ensemble and the second-level ensemble may generate *normal cell* as output, the confidences of the judgements are quite different. The normal cells judged by the first-level ensemble do not need further process but those judged by the second-level ensemble require the attention of pathologists in order to decrease the missing diagnosis of lung cancer patients as more as possible.

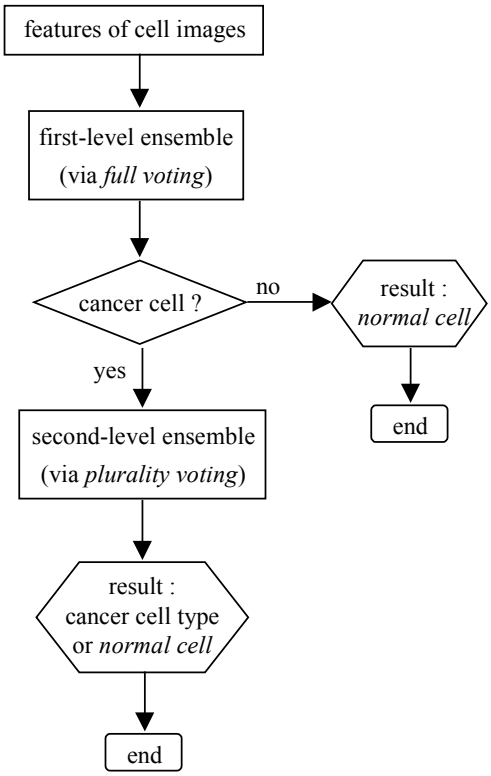

Figure 3. Flowchart of NED

5. Conclusion

Artificial neural networks have already been widely exploited in computer-aided lung cancer diagnosis. However, since there is no rigorous theory indicating how to build a successful artificial neural network based application, the fruits that artificial neural network techniques may produce do not always appear. The artificial neural network ensemble is a recently developed technology, which has the ability of significantly improving the performance of a system where a single artificial neural network is used. Since it is very easy to be used, it has the potential of profiting not only experts in artificial neural network research but also engineers developing real-world applications.

In this paper, we propose an automatic pathological diagnosis procedure named NED, which utilizes artificial neural network ensemble to identify lung cancer cells in the images of the specimens of needle biopsies. The core of NED is a two-level ensemble architecture that is composed of heterogeneous ensembles that not only comprises individual networks with different number of output units but also employs different methods to combine individual predictions. In order to improve the accuracy of false negative identification, we also devised a novel prediction combining method named *full voting* that is utilized in the first-level ensemble. Through adopting those techniques, NED achieves not only high rate of overall identification but also low rate of false negative identification.

NED has been realized in an early stage lung cancer diagnosis system LCDS, which is being transferred into a routine examination following the analysis of X-ray chest films by *Bayi* Hospital currently. In developing the system, we find that employing strong pattern recognition techniques such as artificial neural network ensemble is only one key out of two of improving the performance of the whole system. Only when we get the other key, i.e. strong image processing techniques, we can overcome some obstacles in advancing the performance of the system. In the future, we wish to improve the performance of LCDS in dealing with overlapped cells. Since depth information is ignored by the image capturer, we believe that some kind of 3-D reconstruction is required

to build a tridimensional scene so that those overlapped cells can be distinguished. Moreover, we are glad to apply our two-level ensemble architecture and *full voting* method to more real-world domains.

Acknowledgements

The authors wish to thank their collaborators at *Bayi* Hospital for their suggestive discussion and impressive work on the front-end processing in LCDS system. The comments and suggestions from the anonymous reviewers greatly improve this paper. The National Natural Science Foundation of P.R.China and the Natural Science Foundation of Jiangsu Province, P.R.China, supported this research.

References

[1] Breiman L. Bagging predictors. *Machine Learning* 1996; 24(2): 123-140.

[2] Castleman K R. *Digital image processing*. NJ: Prentice Hall, 1996.

[3] Chan L-W. Weighted least square ensemble networks. In: *Proceedings of the IEEE International Joint Conference on Neural Networks*, 1999. p.1393-1396.

[4] Cherkauer KJ. Human expert level performance on a scientific image analysis task by a system using combined artificial neural networks. In: *Proceedings of the 13th AAAI Workshop on Integrating Multiple Learned Models for Improving and Scaling Machine Learning Algorithms*, 1996. p.15-21.

[5] Chiou YSP, Lure YMF, Ligomenides PA. Neural network image analysis and classification in hybrid lung nodule detection (HLND) system. In: *Proceedings of the IEEE-SP Workshop on Neural Networks for Signal Processing*, 1993. p.517-526.

[6] Cooper LN. Hybrid neural network architectures: equilibrium systems that pay attention. In: Mammone RJ, Zeevi YY editors. *Neural Networks: Theory and Applications*, San Diego, CA: Academic Press, 1991. p.81-96.

[7] Cunningham P, Carney J, Jacob S. Stability problems with artificial neural networks and the ensemble solution. *Artif Intell Med* 2000; 20(3): 217-225.

[8] Gutta S, Wechsler H. Face recognition using hybrid classifier systems. In: *Proceedings of the IEEE International Conference on Neural Networks*, 1996. p.1017-1022.

[9] Hansen LK, Salamon P. Neural network ensembles. *IEEE Trans. Pattern Analysis and Machine Intelligence* 1990; 12(10): 993-1001.

[10] Hansen LK, Liisberg L, Salamon P. Ensemble methods for handwritten digit recognition. In: *Proceedings of the IEEE-SP Workshop on Neural Networks for Signal Processing*, 1992. p.333-342.

[11] Hayashibe R, Asano N, Hirohata H, Okumura K, Kondo S, Handa S, Takizawa M, Sone S, Oshita S. An automatic lung cancer detection from X-ray images obtained through yearly serial mass survey. In: *Proceedings of the International Conference on Image Processing*, 1996. vol.1, p.343-346.

[12] Hornik KM, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. *Neural Networks*, 1989; 2(2): 359-366.

[13] Huang FJ, Zhou ZH, Zhang HJ, Chen TH. Pose invariant face recognition. In: *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition*, 2000. p.245-250.

[14] Kanazawa K, Kubo M, Niki N. Computer aided diagnosis system for lung cancer based on helical CT images. In: *Proceedings of the 13th International Conference on Pattern Recognition*, 1996. vol.3, p.381-385.

[15] Krogh A, Vedelsby J. Neural network ensembles, cross validation, and active learning. In: Tesauro G, Touretzky D, Leen T editors. *Advances in Neural Information Processing Systems 7*, Cambridge, MA: MIT Press, 1995. p.231-238.

[16] Lin JS, Lo SCB, Hasegawa A, Freedman MT, Mun SK. Reduction of false positives in lung nodule detection using a two-level neural classification. *IEEE Trans. Medical Imaging* 1995; 15(2): 206-217.

[17] Liu Y, Yao X. Ensemble learning via negative correlation. *Neural Networks* 1999; 12(10): 1399-1404.

[18] Maclin R, Shavlik JW. Combining the predictions of multiple classifiers: Using competitive learning to initialize neural networks. In: *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, 1995. p.524-530.

[19] Mao J. A case study on bagging, boosting and basic ensembles of neural networks for OCR. In: *Proceedings of the IEEE International Joint Conference on Neural Networks*, 1998. vol.3, p.1828-1833.

[20] Mori K, Hasegawa J, Toriwaki J, Anno H, Katada K. Recognition of bronchus in three-dimensional X-ray CT images with applications to virtualized bronchoscopy system. In: *Proceedings of the 13th International Conference on Pattern Recognition*, 1996. vol.3, p.528-532.

[21] Opitz D, Maclin R. Popular ensemble methods: an empirical study. *J Artif Intell Res* 1999; 11: 169-198.

[22] Penedo MG, Carreira MJ, Mosquera A, Cabello D. Computer-aided diagnosis: a neural-network-based approach to lung nodule detection. *IEEE Trans. Medical Imaging* 1998; 17(6): 872-880.

[23] Schapire RE. The strength of weak learnability. *Machine Learning* 1990; 5(2): 197-227.

[24] Sharkey D editor. *Combining artificial neural nets: Ensemble and modular multi-net systems*. London: Springer-Verlag, 1999.

[25] Sharkey AJC, Sharkey NE, Cross SS. Adapting an ensemble approach for the diagnosis of breast cancer. In: *Proceedings of the International Conference on Artificial Neural Networks*, 1998. p.281-286.

[26] Shimshoni Y, Intrator N. Classification of seismic signals by integrating ensembles of neural networks. *IEEE Trans. Signal Processing* 1998; 46(5): 1194-1201.

[27] Sollich P, Krogh A. Learning with ensembles: how over-fitting can be useful. In: Touretzky D, Mozer M, Hasselmo M editors. *Advances in Neural Information Processing Systems 8*, Cambridge, MA: MIT Press, 1996. p.190-196.

[28] Taniguchi M, Tresp V. Averaging regularized estimators. *Neural Computation* 1997; 9(5): 1163-1178.

[29] Zhou ZH, Chen SF, Chen ZQ. FANNC: A fast adaptive neural network classifier. *Knowledge and Information Systems* 2000; 2(1): 115-129.