DETECTION OF PRIMARY AND SECONDARY CANCERS USING RAMAN SPECTROSCOPY AND SELF-CONSTRUCTING NEURAL NETWORKS**

Zohreh Dehghani-Bidgoli [1*], Tahereh Khamechian [2]

[1] Department of Biomedical Engineering, Kashan Branch, Islamic Azad University, Kashan, Iran; e-mail: dehghani_zohreh@yahoo.com
[2] Department of Medical Pathology, Kashan University of Medical Sciences, Kashan, Iran

The present study aimed to propose a new method for the optimization of neural networks, known as self-constructing neural network (SCNN), to discriminate the Raman spectra of normal tissues, as well as primary and metastatic (secondary) cancers. According to the results, this novel method could significantly improve the ability of the neural network and thoroughly classify the Raman spectra relating to the pathologic states (100% accuracy).

Keywords: *artificial neural networks, Raman spectroscopy, cancer, metastasis, lymph node.*

ОБНАРУЖЕНИЕ ПЕРВИЧНОГО И ВТОРИЧНОГО РАКА С ПОМОЩЬЮ КР СПЕКТРОСКОПИИ И САМОНАСТРАИВАЮЩИХСЯ НЕЙРОННЫХ СЕТЕЙ

Z. Dehghani-Bidgoli [1*], T. Khamechian [2]

УДК 535.375.5;616-006.6

[1] Кашанское отделение Исламского университета Азад, Кашан, Иран; e-mail: dehghani_zohreh@yahoo.com
[2] Кашанский медицинский университет, Кашан, Иран

(Поступила 27 февраля 2018)

Предложен метод самонастраивающейся нейронной сети для различения спектров комбинационного рассеяния нормальных тканей и тканей, пораженных первичным и метастатическим (вторичным) раком. Метод имеет практически 100 % точность и может значительно улучшить способность нейронной сети классифицировать спектры комбинационного рассеяния, относящиеся к патологическим состояниям.

Ключевые слова: *искусственные нейронные сети, спектроскопия комбинационного рассеяния света, рак, метастазирование, лимфатический узел.*

Introduction. Metastasis is the process through which initial tumor cells migrate to a remote or secondary site in the body of the host. Metastasis is a hallmark of malignant tumors as distinguished from benign tumors, which occurs via lymphatic or hematogenous spread. The nearby lymph nodes are primarily involved in metastasis and are known as positive nodes. A common medical practice in this regard is to test the lymph nodes in the vicinity of a tumor site through biopsy during tumor surgery for the examination or removal of the tumor.

Histopathology is considered the “gold standard” for the diagnosis of primary and secondary cancers. The procedure involves the examination of cell morphology using light microscopy in the thin tissue sections stained by hematoxylin and eosin (H&E). Histopathology is associated with several complications, such as invasion (need for biopsy), time-consuming process, and subjective results. Therefore, researchers have proposed noninvasive optical examination methods combined with machine learning and artificial intelligence (AI) algorithms for relevant decision-making processes. Among various optical tissue examination

** Full text is published in JAS V. 86, No. 3 (http://springer.com/10812) and in electronic version of ZhPS V. 86, No. 3 (http://www.elibrary.ru/title_about.asp?id=7318; sales@elibrary.ru).

methods, Raman spectroscopy has the advantage of providing information-rich spectra as the fingerprint of the molecular composition of tissues without alterations or adverse effects on tissues or need for preparation [1, 2].

To date, several studies have focused on the diagnosis of different types of cancer by Raman spectroscopy using various discriminating approaches [3]. The present study aimed to distinguish metastatic (secondary) and primary cancers from normal tissues using Raman spectroscopy and neural networks. Furthermore, to improve the diagnostic accuracy, a novel method has been proposed for the optimization of neural network architectures, known as self-constructing neural networks (SCNN).

Materials and methods. The study was conducted on three patients with metastatic gastric, colon, and breast cancer to the lymph nodes. Three tissue samples were collected from the normal tissues of the primary site, the cancerous region of the primary site, and metastasis to the nearby lymph node (secondary cancer) in each patient. In total, nine samples (three samples from each patient) were obtained. The collected samples were sent to a Raman spectroscopy laboratory, and three spectra were measured in different segments of each sample. The spectra were measured using a micro Raman spectrometer with a 785-nm diode laser. In total, 27 spectra were collected within the wave number range of 500–3500 cm^{-1} (resolution 0.5 cm^{-1}) and were subsequently processed using MATLAB software. Following spectroscopy, diagnosis was performed for each sample at the histopathology laboratory of Shahid Beheshti Hospital in Kashan, Iran.

Preprocessing was carried out to eliminate fluorescence contribution in the spectra. Initially, the fluorescence background was removed using a range-independent background subtraction algorithm (RIA) [4] in the full range of the spectra with an FWHM of 40 cm^{-1}. Then, the noise of the spectra was removed using a zero-order Savitzky–Golay (SG) smoothing filter with the span of 100 spectral points. Finally, the intensity of each spectrum was normalized between zero and one.

Tissue types were classified into three categories of normal, primary, and secondary cancer using a regular multilayer perceptron (MLP) neural network, an architecturally optimized neural network (SCNN), and a completely connected neural network (CCNN) with leave-one-out cross-validation (LOOCV). MLP and CCNN were applied with one hidden layer containing 10 and 30 hidden neurons, respectively. In the SCNN, the number of the hidden neurons and connections between the layers were optimized in each training set to ensure the highest accuracy in classification using a genetic algorithm (GA).

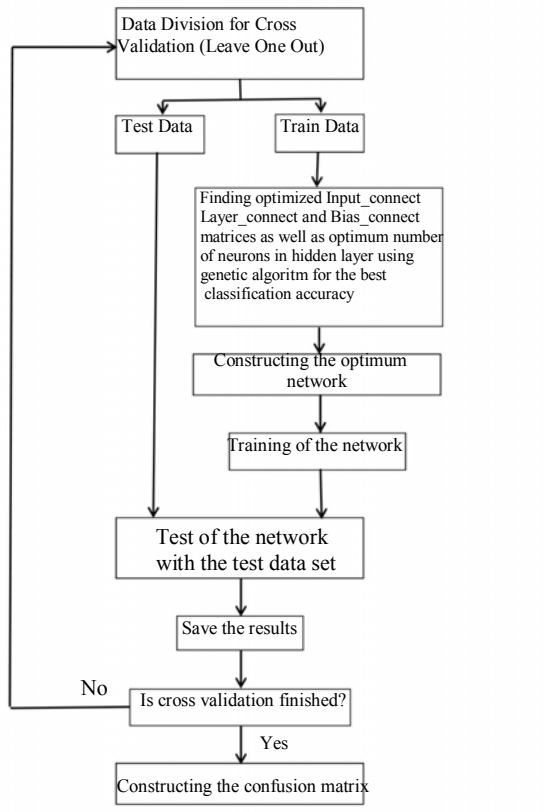

Fig. 1. Diagram of SCNN formation with cross-validation.

The procedure of the SCNN function is depicted in Fig. 1. As shown in the diagram, the optimal Input_connect, Layer_connect, and Bias_connect matrices were obtained for each training set in the cross-validation. These binary matrices (elements: $n \times m$, $n \times n$, and $n \times 1$, respectively) determined the presence or absence of the connections between the m inputs and n layers, connections within the n layers, and bias of the n layers, respectively. In the present study, we assumed $m = 1$ (one input layer) and $n = 2$ (two layers, one hidden, and one output layer). As a result, eight optimum binary elements corresponding to the elements of the matrices were acquired using the GA. Moreover, the optimal number of neurons in the single hidden layer was obtained as an integer within the range of 1–30 using the GA.

The accuracy of classification on the training set was considered to be the objective function of the GA, which should be maximized through optimization. After determining the optimal network for each training set, the performance of the network was assessed in the deposited test set, and the final confusion matrix was constructed based on the test results. It is also notable that to prevent the generation of useless networks without input-output connections, forward connections (input-hidden layer and hidden layer-output) were added to each SCNN before verifying their accuracy.

Results and discussion. Comparison of the mean normalized spectra in the normal, primary, and secondary cancer classes is shown in Fig. 2. Figure 3 illustrates the architecture of MLP, a sample of the optimum SCNN on one of the training sets, and CCNN. With regard to the sample SCNN, there are two backward connections and one direct cascade connection from the input layer to the output layer, and the network has 13 neurons in the hidden layer. However, the connections and number of the hidden neurons varied among the SCNNs obtained in each training set.

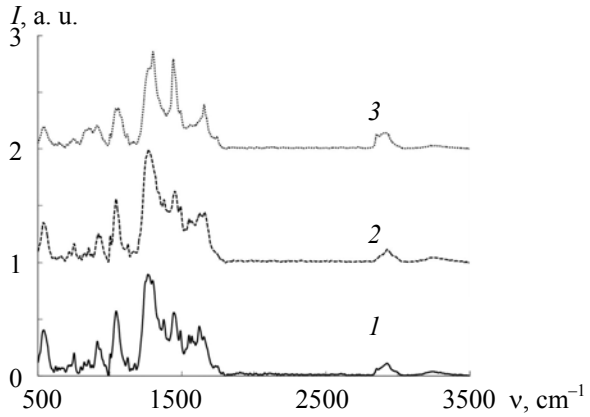

Fig. 2. Mean normalized spectra of normal tissues (1), primary cancers (2), and secondary cancers (3).

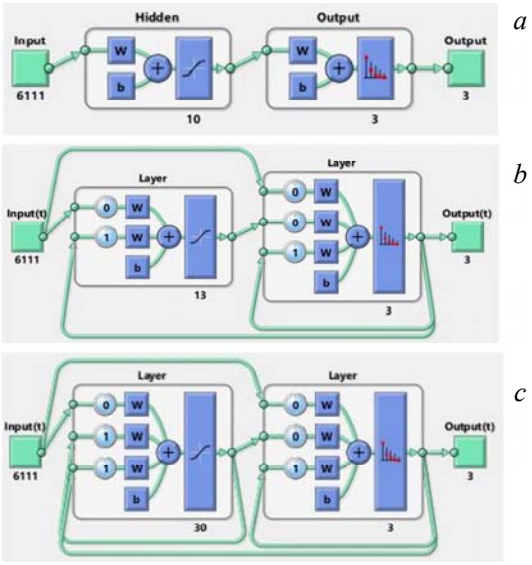

Fig. 3. Architectures of MLP (a), sample of optimized network (SCNN) (b), and CCNN (c).

In order to obtain an overall image of the optimized SCNNs in the current data, the number of each possible connection present in all the SCNNs is shown in Fig. 4. According to the figure, in addition to the manually-added forward connections, there is a direct cascade connection from the input layer to the output layer in all the SCNNs. Furthermore, a backward connection from the output layer to the hidden layer is observed in most of the SCNNs. The average number of hidden neurons in all the SCNNs was 15.

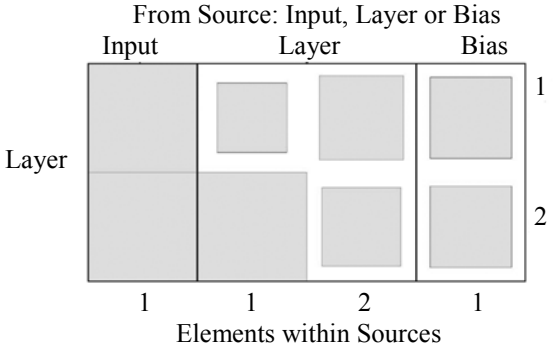

Fig. 4. Superposition of connection matrices of optimized networks obtained in cross-validation (larger square shows the more prominent presence of links in optimized networks).

The confusion matrices of LOOCV-MLP, SCNN, and CCNN classifiers are presented in Tables 1–3. According to these tables, SCNN (100% accuracy) performed prominently superior compared to MLP and CCNN with the accuracy of 59% and 44%, respectively.

TABLE 1. Confusion Matrix of MLP

Target output	N	P	S	
N	7	5	4	0.44
P	0	4	0	1.00
S	2	0	5	0.71
	0.78	0.44	0.55	0.59

TABLE 2. Confusion Matrix of SCNN

Target output	N	P	S	
N	9	0	0	1.00
P	0	9	0	1.00
S	0	0	9	1.00
	1.00	1.00	1.00	1.00

TABLE 3. Confusion Matrix of CCNN

Target output	N	P	S	
N	2	3	1	0.33
P	6	6	4	0.38
S	1	0	4	0.8
	0.22	0.66	0.44	0.44

In the present study, a new algorithm known as SCNN was proposed to determine the optimal neural network to discriminate the Raman spectra of normal tissues from primary and secondary cancers. According to the findings, SCNN could thoroughly classify the Raman spectra in this regard. Furthermore, the pro-

posed approach had favorable adaption with the data and could discriminate cancerous and metastatic spectra of various tissues by concentrating on the cancerous and metastatic differences rather than the inter-tissue differences.

The results of the present study show that the performance of SCNN was superior to that of common MLP and CCNN. This finding indicated that the adequate use of AI techniques and Raman spectroscopy would enable the diagnosis of primary and secondary cancers without the need to interpret complicated Raman spectra. In addition, it seems that the application of SCNN and Raman spectroscopy could contribute to determining the origin of secondary cancers; however, further investigation is recommended in this regard.

Conclusion. The present study aimed to propose a novel method for the optimization of neural network architectures based on the data, which is known as SCNN. The approach could also be used to discriminate the Raman spectra of primary and secondary cancers. According to the results, SCNN could accurately classify all the Raman spectra.

REFERENCES

1. I. P. Santos, E. M. Barroso, T. C. Bakker Schut, P. J. Caspers, C. G. F. van Lanschot, D.-H. Choi, M. F. van der Kamp, R. W. H. Smits, R. van Doorn, R. M. Verdijk, V. Noordhoek Hegt, J. H. von der Thüsen, C. H. M. van Deurzen, L. B. Koppert, G. J. L. H. van Leenders, P. C. Ewing-Graham, H. C. van Doorn, C. M. F. Dirven, M. B. Busstra, J. Hardillo, A. Sewnaik, I. ten Hove, H. Mast, D. A. Monserez, C. Meeuwis, T. Nijsten, E. B. Wolvius, R. J. Baatenburg de Jong, G. J. Puppels, S. Koljenović, *Analyst*, **142**, No. 17, 3025–3047 (2017).
2. L. A. Austin, S. Osseiran, C. L. Evans, *Analyst,* **141**, 476–503 (2016).
3. Q. Tu, C. Chang, *Nanomedicine: Nanotechnol., Biol. Med.*, **8**, No. 5, 545–558 (2012).
4. H. Krishna, S. K. Majumder, P. K. Gupta, *J. Raman Spectrosc.*, **43**, 1884–1894 (2012).