

Raman Spectroscopy-based Breast Cancer Detection Using Self-Constructing Neural Networks

Maliheh Eshraghi-Arani¹, Zohreh Dehghani-Bidgoli^{2*}

1. Department of Computer Engineering, Kashan Branch, Islamic Azad University, Kashan, Iran
2. Department of Biomedical Engineering, Kashan Branch, Islamic Azad University, Kashan, Iran

ARTICLE INFO	ABSTRACT
Article type: Original Paper	Introduction: Accurate and early diagnosis of cancer is an important issue in modern healthcare systems. Raman spectroscopy, as a non-invasive optical technique for evaluating intact tissues at a molecular level, has attracted the researchers' attention. Despite recent advances, efforts are still being made to improve the sensitivity and specificity of Raman spectroscopy-based cancer detection. The present study aimed to identify three classes of breast tissues, that is, normal tissues, benign lesions, and cancer tissues, using an artificial neural network (ANN).
Article history: Received: Nov 10, 2019 Accepted: Apr 26, 2020	Material and Methods: To improve the ANN discrimination power, a novel topologically optimized ANN, known as self-constructing neural network (SCNN), was developed in this study. The ant colony optimization algorithm was applied to optimize the topology of the network. The results of SCNN were compared with the conventional ANN, that is, multilayer perceptron (MLP).
Keywords: Artificial Neural Network Multilayer Perceptron Self-Constructing Neural Network Raman Spectroscopy Breast Cancer	Results: Based on the results, the developed SCNN showed a classification accuracy of 95%. Conclusion: In this study, a novel neural network (SCNN) was proposed, which was topologically optimized to improve the discrimination power of ANNs. The SCNN accuracy was determined to be 95% in Raman spectroscopy-based breast cancer diagnosis.

► Please cite this article as:
Eshraghi-Arani M, Dehghani-Bidgoli Z. Raman Spectroscopy-based Breast Cancer Detection Using Self-Constructing Neural Networks. Iran J Med Phys 2021; 18: 89-95. 10.22038/ijmp.2020.44367.1678.

Introduction

Cancers are considered as severe threats to human health in the modern era. Breast cancer is the most common type of cancer among women, accounting for thousands of deaths each year. Early cancer diagnosis can significantly improve the treatment outcomes [1]. Breast cancer is often characterized by screening mammography, followed by histopathological analysis. Breast-conserving surgery is also prescribed as part of treatment in some cases. The term "breast-conserving surgery" refers to the preservation of healthy tissues as far as possible, while thoroughly removing the tumor. The use of efficient tools for examining large tissue areas and accurate discrimination of lesions in real time can be of substantial importance for surgeons [2].

Extensive research has focused on developing more rapid and accurate methods of cancer diagnosis. Despite major advances, histopathology remains the gold standard diagnostic method. However, histopathology is associated with limitations, such as invasiveness, prolonged response time, and dependence on the experience and skills of pathologists. To overcome these shortcomings, numerous optical techniques have been employed in cancer diagnosis, including optical coherence tomography (OCT), white light reflectance (WLR)

imaging, autofluorescence imaging, and Raman spectroscopy. Due to the lack of molecular data in OCT and WLR, these methods show intrinsically low specificities; similarly, autofluorescence imaging has low specificity [2]. On the other hand, Raman spectroscopy, by analyzing the vibrations of molecular bonds, can provide high molecular specificity.

Raman spectroscopy is a non-invasive optical technique, which assesses intact tissues within a short period and even allows for real-time assessments. Accordingly, this technique has attracted the attention of many cancer researchers [2-12]. Raman spectroscopy is based on the inelastic scattering of monochromatic light. Depending on the inelastic scattering, the color (wavelength) of scattered light differs from the incident light. By measuring the intensity of scattered light at higher wavelengths than the incident wavelength, the Raman spectrum represents a fingerprint of the sample [13], and various lesions (e.g., cancer) can be identified through the analysis of the Raman spectra.

Despite the feasibility of non-invasive, real-time, and *in-vivo* diagnosis of cancer based on the Raman spectra, improvement of the sensitivity and specificity of cancer diagnosis remains a major issue in the clinical application of this technique. Accordingly,

numerous artificial intelligence algorithms have been employed to extract beneficial data from the Raman spectra, including artificial neural networks (ANNs) [14, 15]. In this regard, M. Jermyn et al. used a three-layer neural network for detecting brain cancer in the presence of room light [14]. In another study, J. Liu et al. [15] used convolutional neural networks for identifying chemical species and found that convolutional neural networks perform better than some classification methods, without the need for the preprocessing steps.

In the present study, identification of breast tissues was performed in three classes of normal tissues, benign lesions, and cancer tissues, using an ANN. To improve the ANN discrimination power, a novel ANN, called the self-constructing neural network (SCNN), was developed. The results of SCNN were compared with the conventional ANN, known as multilayer perceptron (MLP). We used SCNN in another study to analyze the Raman spectra of secondary and primary cancers [16]; the full spectra of various primary and secondary cancers were also applied in the network. According to the obtained results, SCNN performed better than MLP. In the mentioned study, genetic algorithm was also used to optimize the neural network architecture, while in the present study, we used the ant colony optimization (ACO) for the same purpose, as ACO has a shorter processing time than other evolutionary-based optimization methods [17, 18]. In other words, in the current study, we studied another version of SCNN for evaluating the Raman spectra of breast cancer.

Materials and Methods

Specimens

A total of 11 breast tissue specimens were obtained from the histopathology laboratory of Shahid Beheshti Hospital in Kashan, Iran, including three normal tissues (tumor margin), five benign lesions (fibrocystic changes), and three cancer tissues (invasive ductal carcinoma). The specimens were fixed in a formalin solution (10% neutral buffered formaldehyde in water).

Spectroscopy

The specimens were examined, using Raman spectroscopy after being removed from the formalin solution. To prevent the interference of the formalin spectrum with the spectra of the specimens, each specimen was exposed to fresh air for several minutes so

that formalin would evaporate from its surface. The spectra were measured within the range of 500-3200 cm^{-1} with a resolution less than 3 cm^{-1} , using a micro-Raman system, equipped with a lens (50x) and a diode laser (785 nm; power, 10 mW). A total of 3-6 spectra were measured for each specimen, and 49 spectra were finally collected. After spectroscopy, the specimens were placed in the formalin solution again and sent back to the histopathology laboratory to continue the histopathological analysis and make the final diagnosis. The spectra of each specimen were assigned a class label, based on the pathologist's diagnosis. Finally, a dataset, containing 14 normal, 18 benign, and 17 cancerous spectra, was obtained.

Preprocessing

For a more accurate detection of the peaks, the resolution of the spectra was initially enhanced to 1 cm^{-1} , using spline interpolation. Following that, the baseline of the spectra (mainly originating from fluorescence) was eliminated, using the algorithm developed by Krishna (2012), known as the range-independent algorithm (RIA) [19]. The RIA cuts the spectrum into determined wavenumber ranges, thereby extrapolating the spectrum on both ends by the linear least square fitting. Next, the extrapolated spectrum is extended by adding two Gaussian peaks with determined heights and widths to both ends, followed by iterative smoothing. In each iteration, the minimum smoothed and original spectra are preserved until complete retrieval of the two terminal Gaussian peaks [19]. In the present study, RIA was used in the full wavenumber range. The height and full width at half maximum (FWHM) of the two added Gaussian peaks were found to be twice the maximum height of the spectrum and 40 cm^{-1} , respectively.

A zero-order Savitzky-Golay filter, with a span of 20 spectral points, was used for smoothing. Following that, the spectra were normalized in terms of the intensity of Amide I band (1,655 cm^{-1}). It should be noted that the ratio of other Raman bands to the Amide I band has been used as a discriminatory feature in several studies on Raman spectroscopy-based cancer diagnosis [3]; it also appeared in all spectra in the current study. Finally, 12 most important bands of biological samples were determined (Table 1), and the normalized intensity of these bands was extracted as a discriminatory feature. All processing was performed in MATLAB 2014a software platform.

Table 1. Position and assignment of major peaks in the Raman spectra of specimens [20-27]

#	Peak position	Assignment	#	Peak position	Assignment
1	830-880	Polysaccharide	7	1110-1130	$\nu_s(\text{C-C})$ skeletal lipids
2	910-940	$\nu(\text{C-C})$ α -helix proline/ valine (protein)	8	1262-1278	Amide III protein/ $\delta_{ip}(\text{C-H})$ lipids
3	1005-1030	$\nu_s(\text{C-C})$ phenylalanine	9	1285-1304	CH_2 twisting and wagging lipids
4	1030-1050	$\delta_{ip}(\text{C-H})$ phenylalanine	10	1422-1442	CH_2 deformation lipids or proteins
5	1050-1068	$\nu_{as}(\text{C-C})$ skeletal lipids	11	1445-1460	$\delta(\text{CH}_2)$ and $\delta(\text{CH}_3)$ proteins
6	1075-1087	$\nu(\text{C-C})$ or $\nu(\text{C-O})$ lipids/ $\nu(\text{C-C})$ or $\nu_s(\text{PO}^2)$ nucleic acids	12	1740-1750	$\nu(\text{C=O})$ lipids

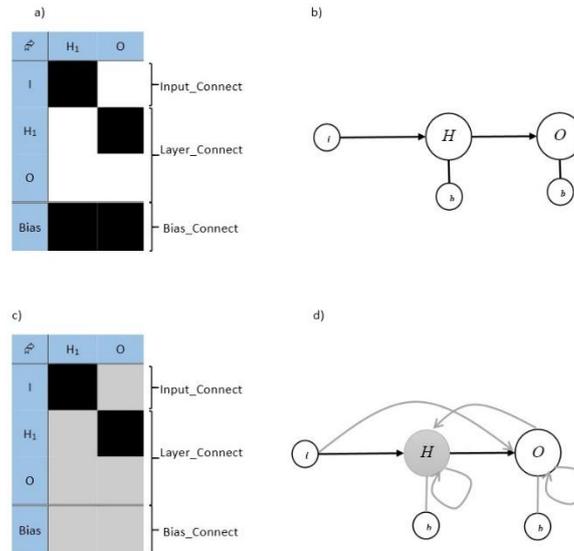


Figure 1. a) Hinton matrix; b) connection diagram of MLP; c) Hinton matrix; and d) connection diagram of SCNN (non-deterministic parameters are shown in gray).

Identification based on neural networks

The spectra were classified using the conventional MLP neural network and SCNN as a novel topologically optimized neural network. The MLP network was used with ten hidden neurons in a single hidden layer with a tangent sigmoid transfer function; three output neurons (per three classes) with a softmax transfer function; and 12 inputs for every feature. The performance and learning functions of the network were selected as cross entropy and scaled conjugate gradient, respectively.

The SCNN optimizes its topology and weight during learning processes. In the present study, the network topology referred to the connections of layers and the size of hidden layers. Also, a similar SCNN to MLP was employed with a single hidden layer. The other specifications of the network (e.g., transfer and learning functions) were also the same as MLP. Figures 1a and 1c depict the Hinton matrices (matrices of network connections) of MLP and SCNN, respectively. The Hinton matrix is composed of three matrices, including the input-to-layer connections (Input-Connect), layer connections to each other (Layer-Connect), and bias-to-layer connections (Bias-Connect).

As shown in Figure 1, the Hinton matrix of MLP included a feed-forward deterministic network (presence/absence of a connection illustrated with dark/white cells, respectively), while the Hinton matrix of SCNN included six non-deterministic elements (gray cells), which could be either zero (absence of connection) or one (presence of connection). Overall, an optimization algorithm can determine the values of these non-deterministic binary elements and the number of hidden neurons as a scalar within the range of 1-30.

Figures 1b and 1d show the schematic diagrams of the input, layer, and bias connections in MLP and SCNN, respectively. All of the non-deterministic values are shown in gray.

In the present study, the ACO was used to optimize the topology of the network, according to the process shown in Figure 2. The objective function of topology optimization was considered as maximum classification accuracy in the training data. As shown in Figure 2, in ACO, sets of artificial ants ($nAnt$) search for the best network to discriminate classes in a high-dimensional variable space (six non-deterministic binary elements for connections and a scalar for the number of hidden neurons). Each set of artificial ants corresponds to one subset of variables. The artificial ants communicate through a virtual pheromone, distributed on the variables, with a primary value of τ . The pheromones change dynamically in iterations and reinforce themselves using positive feedback. Simultaneously, the pheromone decays gradually at an evaporation rate of ρ to emboss the best values. The ACO iteratively executes a loop for $MAXit$ iterations to find the best network. In the present study, we used the ACO parameters, as shown in Table 2.

Table 2. Values and definitions of ACO parameters

Parameters	Values
Maximum number of iterations ($Maxit$)	100
Number of ants ($nAnt$)	9
Primary pheromone (τ)	0.5
Exponential weight of pheromone (α)	0.5
Evaporation rate of pheromone (ρ)	0.5

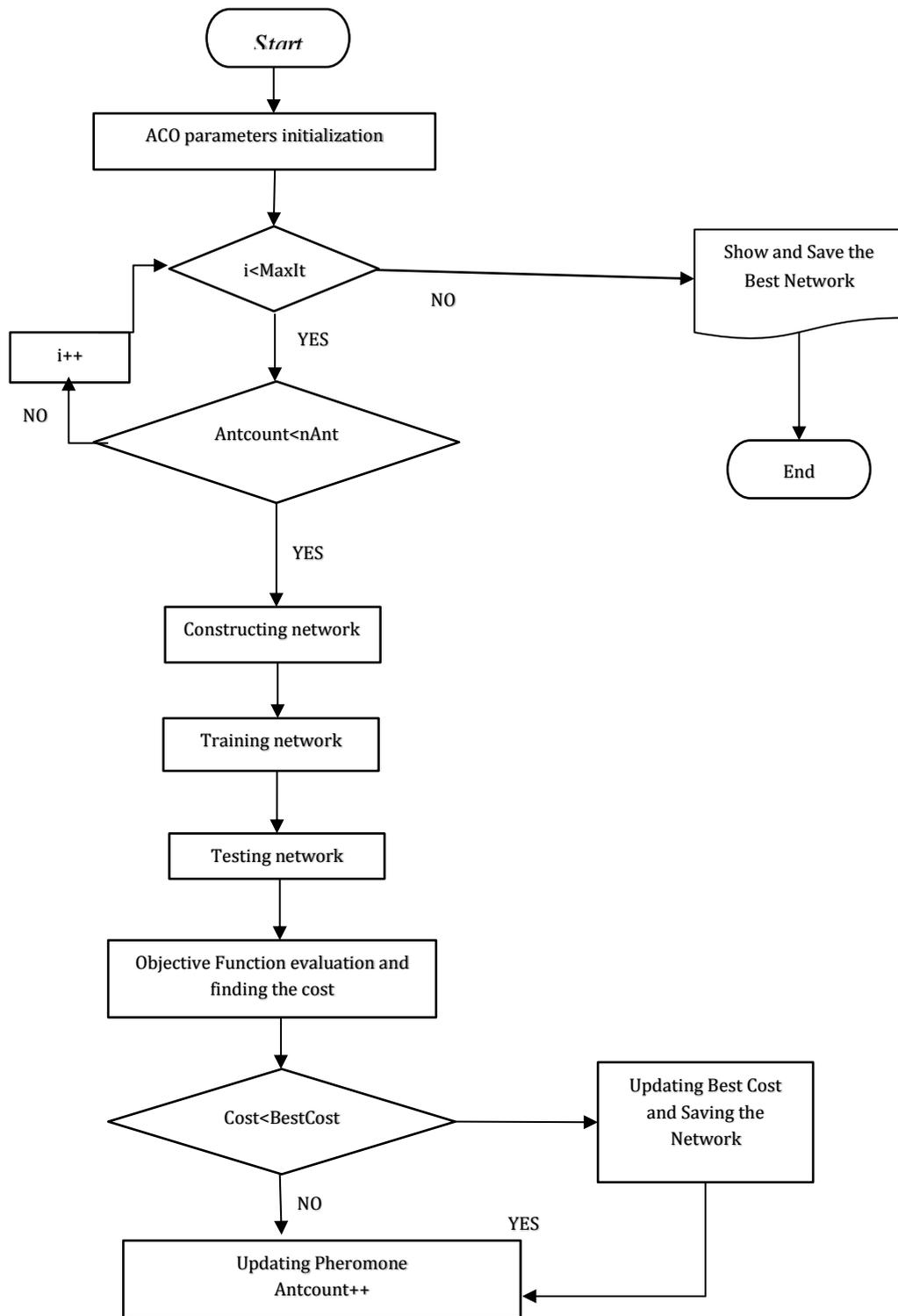


Figure 2. Diagram of developing the optimized neural network in SCNN

To evaluate the generalization of MLP and SCNN networks for the untrained data and to prevent over-fitting, we used the leave-one-out cross-validation (LOOCV). In the LOOCV procedure, a dataset with N instances is trained with N-1 instances and tested with one set-aside instance. Next, the test instance is returned

to the dataset, and the process is repeated with another instance until all instances are included in the test. Finally, a confusion matrix is composed with the N test results. Figure 3 depicts the diagram of the described procedure.

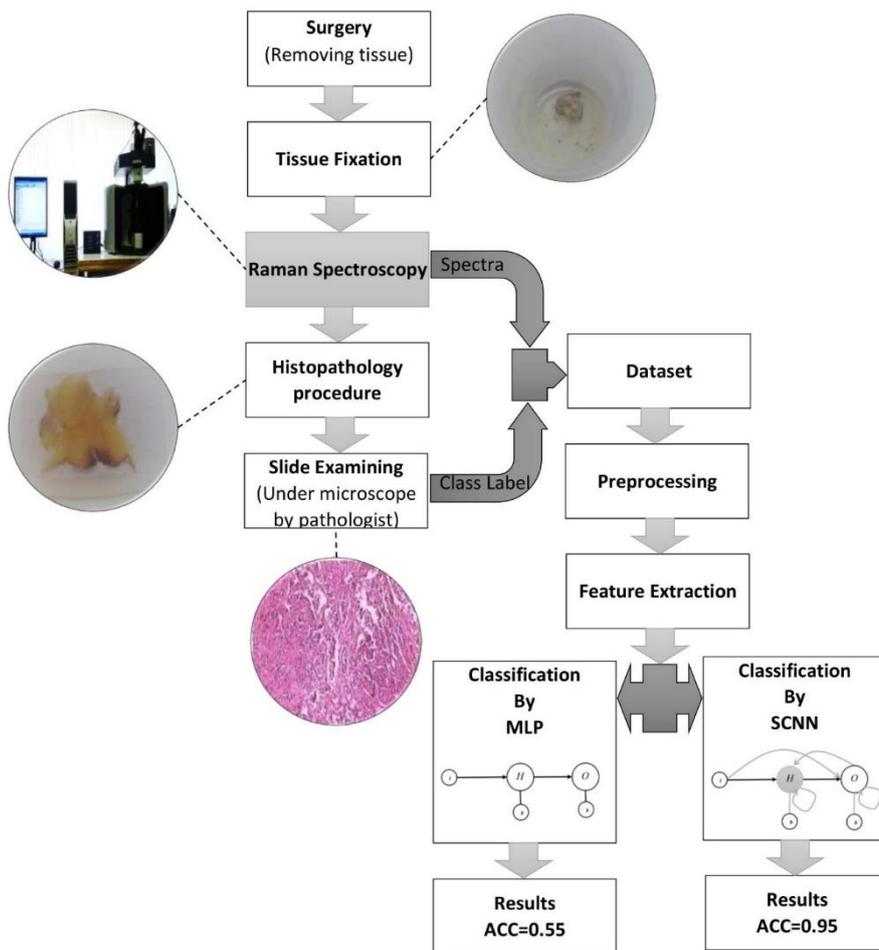


Figure 3. Diagram of the entire procedure

Results

Figure 4 presents the mean pre-processed spectra of the normal, cancerous, and benign tissue classes with solid, dotted, and dashed lines, respectively. Moreover, the spectra are demonstrated as the intensity (vertical axis) versus each wavenumber (horizontal axis).

Figures 5a and 5b depict the topology of the MLP and the optimized SCNN, respectively. The SCNN,

obtained from the entire training dataset, is presented, considering the topology of SCNN changes in each training round of cross-validation.

Tables 3 and 4 show the confusion matrices of MLP and SCNN, respectively. The classification accuracy of MLP and SCNN was estimated at 55% and 95%, respectively.

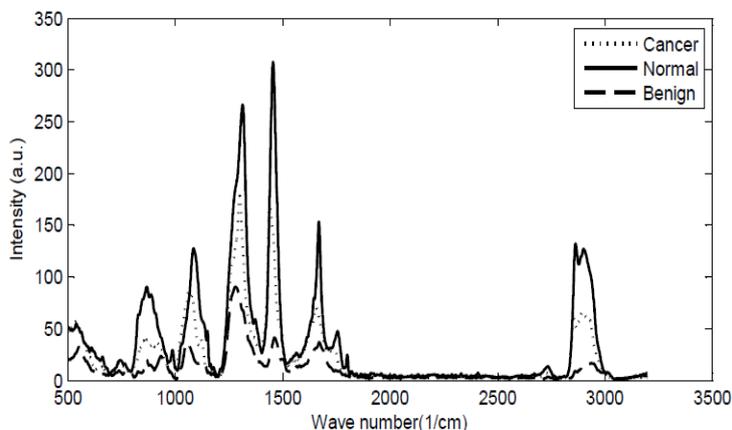


Figure 4. The mean spectra of normal, benign, and cancer tissue classes after preprocessing

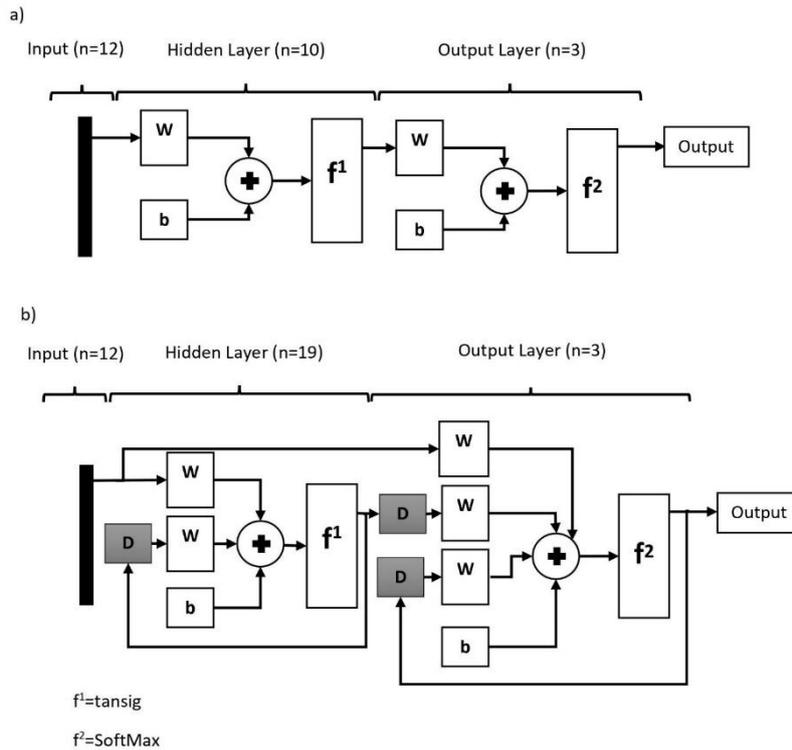


Figure 5. Topology of a) MLP and b) SCNN

Table 3. Confusion matrix of classification based on the MLP

		Target Class			
		A	B	C	S
Output Class	A	12	1	4	0.7059
	B	5	8	1	0.5714
	C	8	3	7	0.3889
	P	0.4800	0.6667	0.5833	0.5510

Table 4. Confusion matrix of classification based on the SCNN

		Target Class			
		A	B	C	S
Output Class	A	17	0	0	1
	B	2	12	0	0.8572
	C	0	0	18	1
	P	0.8947	1	1	0.9592

Discussion

In the current study, we improved the accuracy of Raman spectroscopy-based cancer detection using SCNN, based on MLP. The accuracy of SCNN was determined to be 95% in discriminating between normal, benign, and cancerous breast tissues. Also, we

used the MLP as the basic network, as it is frequently used for analyzing the Raman spectra [6]. The proposed network in our study was based on MLP, which was incorporated into the network (Figure 1).

According to the present findings, SCNN can enhance the power of MLP, although the network complexity and processing time increased, as well. However, the SCNN complexity was significantly lower than deep convolutional neural networks that have been recently employed in the Raman spectrum recognition [15]. Therefore, further comparative investigations on SCNN and deep neural networks are required for reaching a more definite conclusion.

In another study on the same database using the quadratic discriminant analysis, the accuracy of the network was reported to be 73% [20], which is lower than the value obtained in the present study, using SCNN (95%). Also, the mentioned study indicated that the use of proper artificial intelligence approaches for the recognition of Raman spectra can help overcome the limitations of interpreting the Raman spectra and their clinical translation. The current study showed that SCNN could dramatically improve the sensitivity and specificity of Raman spectroscopy-based cancer detection and facilitate its translation to clinics.

Conclusion

In this study, a novel neural network, called the SCNN, was proposed, which was topologically optimized using the ACO on the training dataset. This neural network was used to discriminate the Raman spectra of normal, benign, and cancerous breast tissues.

The results indicated that the SCNN could significantly improve the classification power, based on the MLP.

References

- Richards-Kortum R, Mahadevan-Jansen A, Ramanujam N. Optical spectroscopy vs. the surgical suite [cancer detection]. *IEEE Circuits and Devices Magazine*. 1996 Jul;12(4):34-40.
- Santos IP, Barroso EM, Schut TC, Caspers PJ, van Lanschot CG, Choi DH, et al. Raman spectroscopy for cancer detection and cancer surgery guidance: translation to the clinics. *Analyst*. 2017;142(17):3025-47.
- Mahadevan-Jansen A, Richards-Kortum RR. Raman spectroscopy for the detection of cancers and precancers. *Journal of biomedical optics*. 1996 Jan;1(1):31-70.
- Mahadevan-Jansen A, Richards-Kortum R. Raman spectroscopy for cancer detection: a review. In *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Magnificent Milestones and Emerging Opportunities in Medical Engineering* (Cat. No. 97CH36136). 1997; 6: 2722-8.
- Haka AS, Shafer-Peltier KE, Fitzmaurice M, Crowe J, Dasari RR, Feld MS. Diagnosing breast cancer by using Raman spectroscopy. *Proceedings of the National Academy of Sciences*. 2005 Aug 30;102(35):12371-6.
- [6] Tu Q, Chang C. Diagnostic applications of Raman spectroscopy. *Nanomedicine: Nanotechnology, Biology and Medicine*. 2012 Jul 1;8(5):545-58.
- Kallaway C, Almond LM, Barr H, Wood J, Hutchings J, Kendall C, et al. Advances in the clinical application of Raman spectroscopy for cancer diagnostics. *Photodiagnosis and photodynamic therapy*. 2013 Sep 1;10(3):207-19.
- Kong K, Kendall C, Stone N, Notingher I. Raman spectroscopy for medical diagnostics—From in-vitro biofluid assays to in-vivo cancer detection. *Advanced drug delivery reviews*. 2015 Jul 15;89:121-34.
- Wang W, Zhao J, Short M, Zeng H. Real-time in vivo cancer diagnosis using raman spectroscopy. *Journal of biophotonics*. 2015 Jul;8(7):527-45.
- Austin LA, Osseiran S, Evans CL. Raman technologies in cancer diagnostics. *Analyst*. 2016;141(2):476-503.
- Dehghani-Bidgoli Z, Miran Baygi MH, Kabir E, Malekfar R. Common Raman Spectral Markers among Different Tissues for Cancer Detection. *Iranian Journal of Medical Physics*. 2014;11(4):308-15.
- Dehghani-Bidgoli Z, Baygi MM, Kabir E, Malekfar R. Developing an Instrument-Independent Algorithm for Raman Spectroscopy: A Case of Cancer Detection. *Technology in cancer research & treatment*. 2014 Apr;13(2):119-27.
- Lewis IR, Edwards H. *Handbook of Raman spectroscopy: from the research laboratory to the process line*. CRC press; 2001 Aug 8.
- Jermyn M, Desroches J, Mercier J, Tremblay MA, St-Arnaud K, Guiot MC, et al. Neural networks improve brain cancer detection with Raman spectroscopy in the presence of operating room light artifacts. *Journal of biomedical optics*. 2016 ;21(9):094002.
- Fukuhara M, Fujiwara K, Maruyama Y, Itoh H. Feature visualization of Raman spectrum analysis with deep convolutional neural network. *Analytica chimica acta*. 2019 Dec 9;1087:11-9.
- Dehghani-Bidgoli Z, Khamechian T. Detection of Primary and Secondary Cancers Using Raman Spectroscopy and Self-Constructing Neural Networks. *Journal of Applied Spectroscopy*. 2019 Jul;86(3):528-32.
- Elbeltagi E, Hegazy T, Grierson D. Comparison among five evolutionary-based optimization algorithms. *Advanced engineering informatics*. 2005 Jan 1;19(1):43-53.
- [18] Bergholt MS, Zheng W, Lin K, Ho KY, Teh M, et al. In vivo diagnosis of gastric cancer using Raman endoscopy and ant colony optimization techniques. *International journal of cancer*. 2011 Jun 1;128(11):2673-80.
- Krishna H, Majumder SK, Gupta PK. Range-independent background subtraction algorithm for recovery of Raman spectra of biological tissue. *Journal of Raman Spectroscopy*. 2012 Dec;43(12):1884-94.
- Fallahzadeh O, Dehghani-Bidgoli Z, Assarian M. Raman spectral feature selection using ant colony optimization for breast cancer diagnosis. *Lasers in medical science*. 2018 Nov;33(8):1799-806.
- Shim MG. *Medical Raman spectroscopy, in vivo and ex vivo tissue analysis for cancer diagnosis*. Toronto; 2001.
- Moreno M, Raniero L, Arisawa EA, do Espírito Santo AM, dos Santos EA, Bitar RA, Martin AA. Raman spectroscopy study of breast disease. *Theoretical Chemistry Accounts*. 2010 Mar;125(3):329-34.
- Raniero L, Canevari RA, Ramalho LN, Ramalho FS, dos Santos EA, Bitar RA, Jalkanen KJ, Martinho HS, Martin AA. In and ex vivo breast disease study by Raman spectroscopy. *Theoretical Chemistry Accounts*. 2011 Dec;130(4):1239-47.
- Li S, Chen QY, Zhang YJ, Liu Z, Xiong H, Guo Z, et al. Detection of nasopharyngeal cancer using confocal Raman spectroscopy and genetic algorithm technique. *Journal of biomedical optics*. 2012 Dec;17(12):125003.
- Zhao J, Lui H, McLean DI, Zeng H. Real-time Raman spectroscopy for noninvasive in vivo skin analysis and diagnosis. *New developments in biomedical engineering*. 2010 Jan 1;24:455-74.
- Zhou Y, Liu CH, Sun Y, Pu Y, Boydston-White S, Liu Y, Alfano RR. Human brain cancer studied by resonance Raman spectroscopy. *Journal of biomedical optics*. 2012 Nov;17(11):116021.
- Dehghani-Bidgoli Z, Baygi MM, Kabir E, Malekfar R. A comparative study between carcinoma and sarcoma using Raman spectroscopy. *Journal of Applied Spectroscopy*. 2014 Jan 1;80(6):893-8.