Lung-Nodule Segmentation Using a Convolutional Neural Network with the U-Net Architecture

Vicente Hernández-Solis, Arturo Téllez-Velázquez, Antonio Orantes-Molina, and Raúl Cruz-Barbosa(✉)

Applied Artificial Intelligence Laboratory, Computer Science Institute, Universidad Tecnológica de la Mixteca, 69000 Huajuapan de León, Oaxaca, Mexico
hesv880702ps@ndikandi.utm.mx, {atellezv,tonito,rcruz}@mixteco.utm.mx

Abstract. Lung cancer is one of the types of cancer that claims the most lives globally. For screening purposes, computed tomography scans are the most reliable source for nodule detection, as it reveals the structure of the chest, through a three dimensional representation, in which lung lesions can be fully observed. For early cancer detection, it is necessary to use computed radiography and tomography of the thorax, as well as searches for potentially malignant nodules by specialists. In this paper, lung nodule segmentation was performed using the LIDC IDRI public database, which includes images of computed tomographies, by means of a modified U-Net convolutional neural network. The experimental results have shown that our proposal achieves a Dice similarity coefficient of 88.1% and accuracy of 99.78%, which improves nodule segmentation performance in comparison with other architectures used in the literature.

Keywords: Lung cancer · Image segmentation · Convolutional neural network · U-Net · Computed tomography

1 Introduction

When it comes to lung cancer, an early clinical diagnosis allow specialists to rule out the presence of malignant nodules and can lead to the correct choice of an appropriate treatment to combat this cancer [7].

Among all the existing forms of cancer, lung cancer is the type that causes the most deaths annually [26]. To detect it, an experienced radiologist searches for lung nodules visually by using advanced medical tests such as computed radiography and tomography. This procedure is often a highly complex task, however, even for specialists. The test consists of making use of a series of digital images that are taken transversely in the form of equidistant cross sections (usually 3 mm apart from one another), from which it is possible to get a three-dimensional representation of the patient's lungs [28].

E. Roman-Rangel et al. (Eds.): MCPR 2021, LNCS 12725, pp. 335–344, 2021.
https://doi.org/10.1007/978-3-030-77004-4_32

With the development of deep learning, many convolutional-neural-network methods (CNN) have yielded surprising results in carrying out computed radiography and tomography image-segmentation tasks [17]. For example, some research papers have dealt with image segmentation using fuzzy logic, active contour algorithms, and clustering; but the effectiveness of these techniques has been largely outdone by machine learning and CNN [15,18,21].

This research paper tackles the subject of lung-nodule segmentation, making use of the LIDC-IDRI public database of tomographies and the U-Net CNN model in order to segment potentially malignant nodules in patients at risk of suffering lung cancer.

This article is organized as follows. Section 2 deals with papers that have tackled the subject of lung-nodule segmentation. Section 3 introduces our implementation of CNN, after which our experiment design will be laid out. Section 4, follows, in which the experiment's results are given; and finally, in Sect. 5, our conclusions are offered, and future work is laid out.

2 Related Work

The analysis of computed tomographies for the detection of nodules begins with the segmentation process. This stage is of vital importance, because an incorrect segmentation would result in false clinical diagnoses, which would directly affect patients. Some papers [1] estimate that test data may miss from 5% to 17% of lung nodules due to a segmentation algorithm's shortcomings. Next, we present some segmentation methods from the literature, which are used to segment pulmonary nodules.

One of the most common ways to perform image segmentation is through the region-based approach, such as the one presented in [20], which uses the fast-marching method to separate the image into regions with similar features. In order to segment lung nodules, multiple seed points are required to produce a set of regions on input image. Finally, they are merged by combining regions that grow applying k-means algorithm.

In one case in particular, Nithila and Kumar [18] developed a nodule segmentation technique based on a Fuzzy C-mean algorithm and an active contour model. The reconstruction of the lung parenchyma was performed by a Gaussian filter and the segmentation was completed with a cluster algorithm.

Another paper has proposed a hybrid system which makes use of the following techniques [14]: morphological operation, dot-enhancement filter based on the Hessian matrix, fuzzy-connectedness segmentation, local-density maximum algorithm, geodesic-distance map, and regression tree classification. First, all the potential nodules are generated to classify them among three types: peripheral, those attached to the chest wall, or those on the mediastinum. Finally, by means of a tree structure, the nodules are detected and classified according to their location, size, and shape.

By contrast, [12] employs discriminative random fields (DRF) to segment 3D volumes of lung nodules in computed tomographies using only one seed point per nodule. First, various features are calculated, such as the estimated radius

and approximate segmentation, making use of morphological filtering. Next, a supervised training of the DRF model is conducted to achieve a more accurate general segmentation.

Other less conventional methods as [23], have used the vanilla optical flow method to process the computed tomography slices. To do this, they created an image frame sequence with timeline-ordered CT slices and thereby achieve lung nodule segmentation by computing the detected changes between frames.

In the case of CNN-based segmentation, an interesting proposal has been published in [4], which makes use of a fully CNN (FCN). Nodule detection is then carried out through binary thresholding within the lung area; next, the approximate nodule segmentation is done through the level set method. Finally, nodule segmentation is obtained based on the coordinate system transformation.

Another CNN proposal [2] uses a dual-branch residual network (DB-ResNet), which simultaneously captures intensity features from multiple views and scales of different nodules in CT images and consequently performs image segmentation. Additionally, they propose a central intensity pooling layer (CIP), to extract the intensity features of the central voxel of the block, and subsequently with CNN obtain the convolutional features of the central voxel. Regarding lung-nodule detection, other papers as [11] have tackled the problem of voxel-to-voxel segmentation by means of a multi-view CNN, by which it is possible to determine if a voxel corresponds to a nodule.

3 Convolutional Neural Networks

The CNNs arose from the study of the human eye's visual cortex, and have been used for image recognition since 1980. CNNs are not limited to computer-vision tasks; rather, they are also used in other endeavors, such as voice recognition and natural language processing [8].

A CNN is a supervised learning model that has the ability to identify objects, since it can extract features from their inputs, especially from images. To do this, a CNN is comprised of several specialized hidden layers that have certain hierarchy. This hierarchy is represented by different levels of abstraction, by which the composition of simple or lower-level features allow for the learning about complex or higher-level features. This means that the first layers can detect geometric primitives, and as we go deeper through the layers, they become more and more specialized, so that deeper layers are able to recognize more complex shapes [3].

Figure 1 illustrates the main operations involved in a CNN's processing. For example, Fig. 1a shows the convolution layer, while Fig. 1c depicts the pooling operation.

3.1 U-Net Architecture

The U-Net architecture was initially designed to segment neural structures into sets of images obtained by electronic microscope. However, this architecture can also be used to classify and segment biomedical images [19].

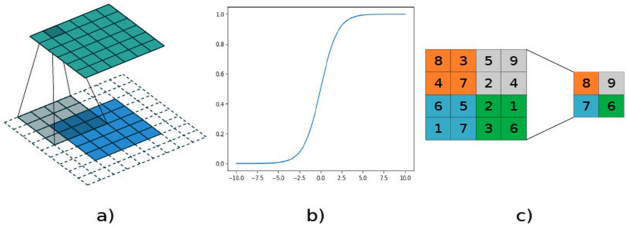

Fig. 1. Basic CNN operations: a) Convolution operation: cross product between a kernel and the image [6]. b) Activation function: it provides non-linearity to the network. c) Pooling operation: it reduces image size while preserving its main features [22]

The U-Net architecture (see Fig. 2), consists of two stages: contraction and expansion. The first stage is a standard CNN, which consists of the repeated application of convolutions with 3×3 filters without padding, which reduces the image's dimensions. Following this, a rectified linear unit layer (ReLU) and a 2×2 max-pooling layer with a stride of 2 is used to reduce the image's dimensions by half, while with each reduction the filter number is doubled, which is to say, one obtains double the feature maps. This stage is mainly used for automated feature extraction from the input images, which allows to improve the segmentation task.

In each step of the expansion stage, a transposed convolution of the feature map and a concatenation with the corresponding feature map cropped from its contraction counterpart are performed. Two convolutions with 3×3 filters and a ReLU layer are then added. Cropping the feature map is necessary, given that in each convolution the image's dimensions are reduced. A convolution with a $1 \times 1 \times n$ filter is used on the last layer to map each 64-component feature vector to the desired total number of classes n [19]. This stage combines both the features obtained from the first stage and the output of the transposed convolution layer to obtain accurate segmentations, although there is a loss of information due to cropping operation which adjusts the dimensions of the output.

The U-Net network is also called a fully convolutional network, since it only uses convolutional layers to perform image segmentation [13]. In other words, it has no fully connected layers, which reduces the number of weights to be estimated [9]. Additionally, this architecture does not require a large number of labeled images for training and it can obtain very accurate segmentations. These characteristics are particularly useful when working with biomedical images since a usual setting is that there are few labeled examples [19].

3.2 U-Net for Nodule Segmentation

In this work, a modified U-Net architecture was proposed to train the network with the LIDC-IDRI public database, as it is shown in Fig. 3. Unlike the original architecture, the input and output images preserve the same size, that is, a resolution of 128×128 pixels is used. This is because both the convolution and the transposed-convolution are carried out with zero padding, which avoids

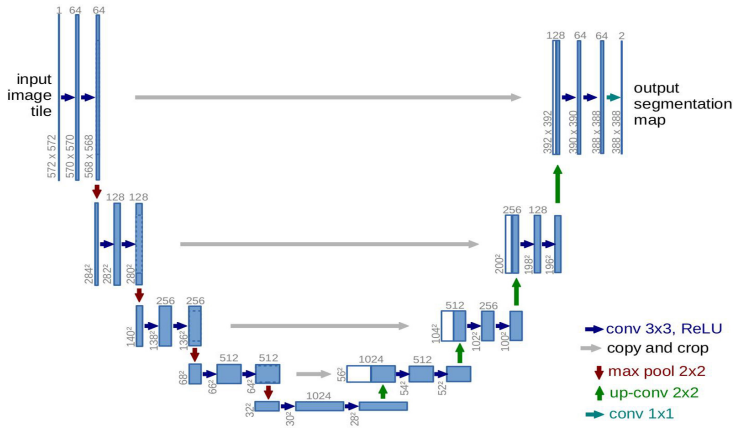

Fig. 2. U-Net deep neural network architecture [19]

that image dimensions have to be reduced in each convolution operation. As a consequence, the concatenation in the second stage (transposed-convolution) is performed without cropping. The number of initial filters is 32 and each time it is doubled (64, 128, and so on), while the image dimensions are reduced by half due to max-pooling layer. In the network's final layer, the segmentation is performed for two classes; in other words, it decides if a pixel belongs to a nodule or not. With our modified U-Net network, there is no information loss in the second stage. Also, the concatenation of the first-stage features and the transposed convolution layer output is direct, thus covering the entire input image.

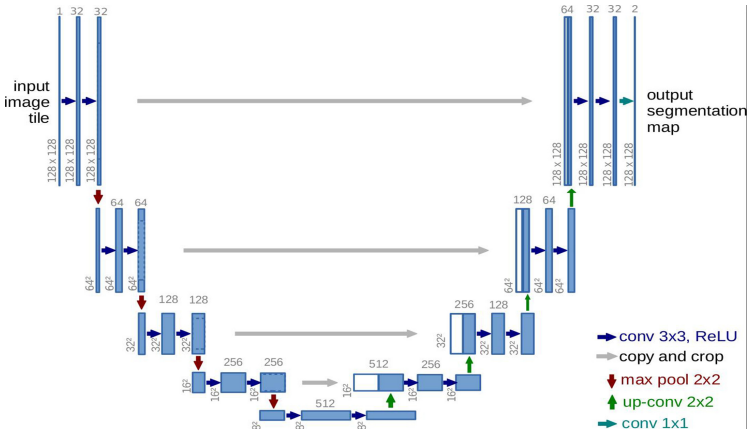

Fig. 3. Modified U-Net deep neural network architecture

Having both the segmentation performed by specialists and that performed by the CNN, the used evaluation criterion is the Dice Similarity Coefficient (DSC), which is given by the following equation [10]:

$$DSC = \frac{2\,|X \cap Y|}{|X| + |Y|} \tag{1}$$

where X is the segmented area by specialists and Y is the segmented area by U-Net. In other words, the DSC is used to measure the overlap between the two regions, which in this case are the nodule-segmentation results. When the DSC equals 1, it is because the segmented regions overlap completely, whereas if the DSC equals 0, there is no overlap between the segmented regions.

4 Results and Discussion

In this section, both the image set and the hardware and software used during implementation are described. Following is an analysis of our results, as well as a comparison with existing results in the literature.

The used image set was obtained from the public database LIDC-IDRI [5]. This database consists of 244,617 images from 1308 studies carried out on 1010 patients. The nodules were identified by as many as four specialists. The selected nodule images were those that were indicated by the four specialists, while the nodules identified by three or fewer specialists were ruled out. In total, there were 900 identified nodules with a diameter of greater than 3 mm.

Additionally, there is information for each study on the patient and on the nodules' properties, such as difficulty of detection, internal structure, calcification, sphericity, margin, degree of lobulation, extension of spiculation, texture, and an evaluation of the probability of malignancy.

To carry out the CNN training, the set of available nodules was divided into training (68%), validation (12%), and test (20%) subsets. The first two subsets were used to obtain the network's best hyperparameters, which turn out to be a learning rate of 0.0005, an initial filter number equal to 32, a batch size of 32, and a filter size of 3. Subsequently, using these hyperparameters, the network is retrained with different sizes of the training subset, in order to measure and compare the performance of the network using the corresponding test subset.

The experiments were conducted on a computer with an Intel(R) Xeon(R) CPU E5-2630 processor, which has 32 processing threads and runs the Ubuntu operating system with 16 GB of RAM memory. In terms of software, the Python programming language was used, along with the TensorFlow library.

Having conducted the corresponding experiments, the results analysis and discussion follow. As we have mentioned, different test-set sizes were used with the goal of making comparisons with other methods. For example, a size of 493 nodules was considered in [25]; also a size of 393 nodules was used in [24]; as well a size of 128 nodules was proposed in [16]. For the purposes of comparing the results with those reported in [16], sizes of 246 and 196 nodules were considered, that is, half of the aforementioned sizes. Table 1 shows the results obtained for

each proposed test-set size, where classification accuracy and DSC coefficient are used as measures of convolutional network performance. Here it can be observed that as the test set decreases the classification accuracy and the DSC coefficient increases. This is expected behavior since the convolutional network learns to discriminate better with more information as the training set increases.

Table 1. Comparison of the performance of the U-Net architecture with different test-set sizes

Test size [Nod]	Accuracy [%]	DSC [%]
493	99.72	84.27
393	99.75	85.62
246	99.76	87.46
196	99.78	88.10

Table 2 presents the performance results of our U-Net proposal compared with other methods [16,18,24,25,27], which also deal with this problem of segmentation using the same database. Observe that the main performance measure for the comparison was DSC, although the test size and the classification accuracy are also included as complementary information. Note that in row 7, column 4 of Table 2, our U-Net proposal reached an average DSC of 88.10% in the test set, which indicates a high degree of similarity between the segmentations obtained by our proposal and those proposed by specialists.

Table 2. Comparison between our proposal (last line) and some other proposal on the state of the art, using the LIDC-IDRI database. The symbol '–' indicates that there is no available information about the accuracy measure or test size in the corresponding articles

Method	Test size [Nod]	Accuracy [%]	DSC [%]
PN-SAMP [27]	–	97.58	74.05
MV-CNN [24]	393	–	77.67
CF-CNN [25]	493	–	82.15
DLGC [16]	128	–	83.00
Fuzzy C-mean [18]	–	98.95	84.00
U-Net	196	99.78	88.10

Comparing the results in Tables 1 and 2, it is observed that the U-Net results are better, in terms of DSC, for all the test set sizes (493, 393, 128) used. That is, the DSC obtained with the U-Net is higher than the publications listed in Table 2, including the article using a small test set (128) compared to the smallest

U-Net set (196). To shed further light on the virtues of our proposal, which allow to improve the original U-Net architecture and related literature results (Table 2), some of the best visual results of the nodule segmentation on the test set are shown below in Fig. 4. For greater clarity, the images are shown in pairs (vertical orientation); the first image is the segmentation performed by the specialists (in green), and the second is the segmentation obtained by the CNN (in red). These segmentation results individually reach over 97% DSC due to the nodules' smooth shape, which is most often circular or oval. It is worth highlighting that the nodule's size also exercises a strong influence, given that the majority of these nodules have a diameter of greater than 4 mm, which aids in their identification.

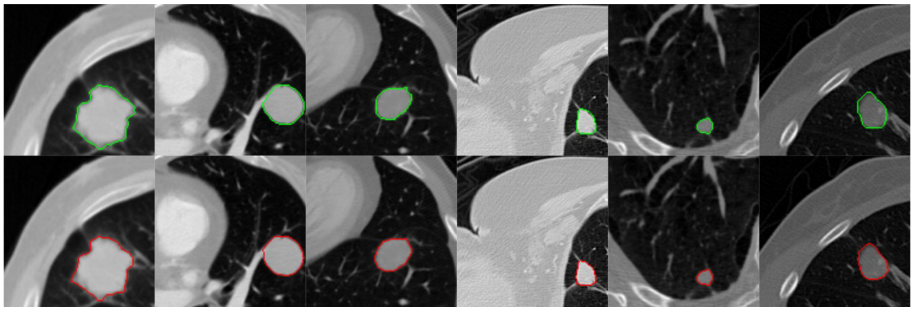

Fig. 4. Regularly shaped nodules segmented by experts in green (above) and nodules segmented by our U-Net proposal in red (below) (Color figure online)

Due to the great variety in lung nodules' shape and size, not all of them can be segmented correctly. For example, Fig. 5 shows some nodules, individually obtaining a DSC of less than 70%, that turned out to be difficult to segment under our proposal, since they did not have a smooth or regular shape. At times, nodules are spiculated or have a diameter close to 3 mm. These factors make their identification highly difficult.

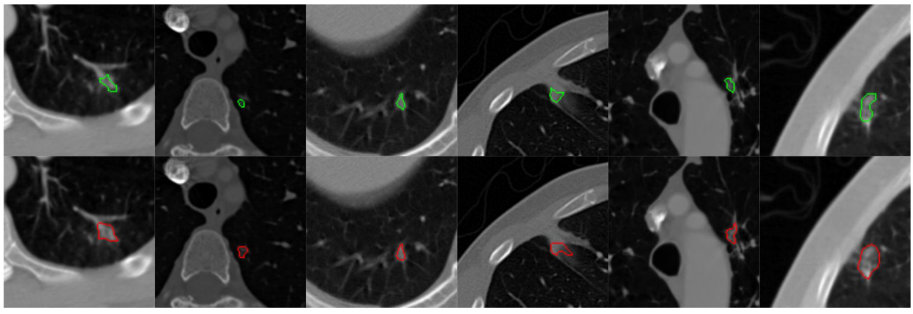

Fig. 5. Irregularly shaped (spiculated) nodules segmented by experts in green (above) and nodules segmented by our U-Net proposal in red (below) (Color figure online)

5 Conclusions

The impact of a modified U-Net deep architecture on the performance of lung-nodule segmentation was presented in this work. The experimental results have shown that the proposed U-Net modification obtains the highest DSC performance (88.10%) for the analyzed dataset, which outperforms other proposals on the state of the art. In spite of the fact that it has problems with performing nodule segmentation on irregular or spiculated shapes, the results are highly competitive.

As future work, several hybrid approaches could be explored with the aim to improve the DSC coefficient and consequently making efficiently progress on lung-nodule detection. For example, a hybrid approach composed of a U-Net-CNN and an adaptive morphological filter could be proposed. Additionally, this proposal should be scaled to distributed computing procedures, thus improving training times and, as a consequence, timely decision-making as well.

References

1. Armato, S.G., Sensakovic, W.F.: Automated lung segmentation for thoracic CT. Acad. Radiol. **11**(9), 1011–1021 (2004)
2. Cao, H., et al.: Dual-branch residual network for lung nodule segmentation. Appl. Soft Comput. **86** (2020)
3. Chollet, F.: Deep learning with Python, 1st. edn. Manning Publications Co. (2018)
4. Chunran, Y., Yuanvuan, W., Yi, G.: Automatic detection and segmentation of lung nodules on CT images. In: 11th International Congress on Image and Signal Processing. BioMedical Engineering and Informatics, pp. 1–6. Beijing, China (2018)
5. CIA: Lung image database consortium and image database resource initiative. https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI. Accessed Nov 2019
6. Dumoulin, V., Visin, F.: A guide to convolution arithmetic for deep learning. ArXiv abs/1603.07285 (2016)
7. Fourcade, A., Khonsari, R.H.: Deep learning in medical image analysis: a third eye for doctors. J. Stomatol. Oral Maxillofacial Surg. **120**, 279–288 (2019)
8. Géron, A.: Hands-On Machine Learning with Scikit-Learn & Tensorflow, 1st edn. O'Reilly Media Inc. (2017)
9. Gulli, A., Kapoor, A., Pal, S.: Deep Learning with TensorFlow 2 and Keras, 2nd edn. Packt Publishing (2019)
10. Havaei, M., et al.: Brain tumor segmentation with deep neural networks. Med. Image Anal. **35**, 18–31 (2017)
11. Litjens, G., et al.: A survey on deep learning in medical image analysis. Med. Image Anal. **42**, 60–88 (2017)
12. Liu, B., Raj, A.: Discriminative random field segmentation of lung nodules in CT studies. Comput. Math. Methods Med. **2013**, 1–9 (2013)
13. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. arXiv preprint arXiv:1411.4038v2 (2015)
14. Lu, L., Tan, Y., Schwartz, L.H., Zhao, B.: Hybrid detection of lung nodules on CT scan images. Med. Phys. **42**(9), 5042–5054 (2015)

15. Manikandan, T., Bharathi, N.: Lung cancer diagnosis from CT images using fuzzy inference system. In: Das, V.V., Thankachan, N. (eds.) CIIT 2011. CCIS, vol. 250, pp. 642–647. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-25734-6_110
16. Mukherjee, S., Huang, X., Bhagalia, R.R.: Lung nodule segmentation using deep learned prior based graph cut. In: 14th IEEE International Symposium on Biomedical Imaging, pp. 1205–1208. Melbourne, VIC, Australia (2017)
17. Ni, J., Wu, J., Tong, J., Chen, Z., Zhao, J.: GC-Net: global context network for medical image segmentation. Comput. Methods Programs Biomed. **190**, 1–10 (2020)
18. Nithila, E.E., Kumar, S.S.: Segmentation of lung nodule in CT data using active contour model and Fuzzy C-mean clustering. Alex. Eng. J. **55**(3), 2583–2588 (2016)
19. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: 18th Medical Image Computing and Computer-Assisted Intervention, pp. 234–241. Munich, Germany (2015)
20. Savic, M., Ma, Y., Ramponi, G., Du, W., Peng, Y.: Lung nodule segmentation with a region-based fast marching method. Sensors **21** (2021)
21. Sivakumar, S., Chandrasekar, C.: Lung nodule segmentation through unsupervised clustering models. ICMOC **38**(2012), 3064–3073 (2012)
22. Srinivas, M.: Max pooling in convolutional neural network and its features. https://analyticsindiamag.com/max-pooling-in-convolutional-neural-network-and-its-features/. Accessed Oct 2020
23. Suji, R.J., Bhadouria, S.S., Dhar, J., Godfrey, W.W.: Optical flow methods for lung nodule segmentation on LIDC-IDRI images. J. Digit. Imaging **33**, 1306–1324 (2020)
24. Wang, S., et al.: A multi-view deep convolutional neural networks for lung nodule segmentation. In: 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 1752–1755. Jeju, Korea (South) (2017)
25. Wang, S., et al.: Central focused convolutional neural networks: developing a data-driven model for lung nodule segmentation. Med. Image Anal. **40**, 172–183 (2017)
26. WHO: Latest global cancer data. https://www.who.int/cancer/PRGlobocanFinal.pdf. Accessed Oct 2019
27. Wu, B., Zhou, Z., Wang, J., Wang, Y.: Joint learning for pulmonary nodule segmentation, attributes and malignancy prediction. In: 15th IEEE International Symposium on Biomedical Imaging, pp. 1109–1113. Washington, D.C., USA (2018)
28. Wu, J., Qian, T.: A survey of pulmonary nodule detection, segmentation and classification in computed tomography with deep learning techniques. J. Med. Artif. Intell. **2**(8), 1–12 (2019)