Original Paper

Assessment of Deep Learning Methods for Differentiating Autoimmune Disorders in Ultrasound Images

CORINA MARIA VASILE[1,2], ANCA LOREDANA UDRIȘTOIU[3], ALICE ELENA GHENEA[4], VLAD PADUREANU[5], ȘTEFAN UDRIȘTOIU[3], LUCIAN GHEORGHE GRUIONU[6], GABRIEL GRUIONU[7], ANDREEA VALENTINA IACOB[3], MIHAELA POPESCU[8]

[1]PhD School Department, University of Medicine and Pharmacy of Craiova, Romania

[2]Department of Pediatric Cardiology, County Clinical Emergency Hospital of Craiova, Romania

[3]Faculty of Automation, Computers and Electronics, University of Craiova, Romania

[4]Department of Bacteriology-Virology-Parasitology, University of Medicine and Pharmacy of Craiova, Romania

[5]Department of Internal Medicine, University of Medicine and Pharmacy of Craiova, Romania

[6]Faculty of Mechanics, University of Craiova, Romania

[7]Department of Medicine, Indiana University School of Medicine, Indianapolis, United States

[8]Department of Endocrinology, University of Medicine and Pharmacy of Craiova, Romania

ABSTRACT: At present, deep learning becomes an important tool in medical image analysis, with good performance in diagnosing, pattern detection, and segmentation. Ultrasound imaging offers an easy and rapid method to detect and diagnose thyroid disorders. With the help of a computer-aided diagnosis (CAD) system based on deep learning, we have the possibility of real-time and non-invasive diagnosing of thyroidal US images. This paper proposed a study based on deep learning with transfer learning for differentiating the thyroidal ultrasound images using image pixels and diagnosis labels as inputs. We trained, assessed, and compared two pre-trained models (VGG-19 and Inception v3) using a dataset of ultrasound images consisting of 2 types of thyroid ultrasound images: autoimmune and normal. The training dataset consisted of 615 thyroid ultrasound images, from which 415 images were diagnosed as autoimmune, and 200 images as normal. The models were assessed using a dataset of 120 images, from which 80 images were diagnosed as autoimmune, and 40 images diagnosed as normal. The two deep learning models obtained very good results, as follows: the pre-trained VGG-19 model obtained 98.60% for the overall test accuracy with an overall specificity of 98.94% and overall sensitivity of 97.97%, while the Inception v3 model obtained 96.4% for the overall test accuracy with an overall specificity of 95.58% and overall sensitivity of 95.58%

KEYWORDS: *Ultrasound imaging, autoimmune disorders; deep learning, convolutional neural networks, transfer learning.*

Introduction

Autoimmune thyroid diseases are thought to be the most prevalent autoimmune disease with a 5% incidence and a gradually growing prevalence in females. Adult women are more likely than men to experience thyroid autoimmunity, and irregular thyroid function is more common in this setting (7%-9% in females vs. 1%-2% in males) [1].

AITDs are T-cell mediated autoimmune disorders caused by organ-specific immune system deregulation. The mechanisms underlying this autoimmune reaction are yet to be fully understood, but an association between genetic predisposition and environmental factors has been shown to initiate the autoimmune process [2].

Thyroiditis diagnosis is not always quick and immediate, since, in the early stages of the condition, precise laboratory test modifications are not detectable. Some characteristics, such as thyroid gland scale, form, and echogenicity, can now be assessed earlier and more accurately as imaging methods advance.

As a consequence, ultrasonography has been useful for evaluating the thyroid gland, especially in uncertain cases where morphological and functional characteristics on color Doppler mapping may indicate the actual enteropathogenic of the disease. Any of these characteristics are generally accepted in the global literature [3].

Due to its safety, lower costs, and portability, ultrasound imaging diagnosis became an important technique used for visualizing and

10.12865/CHSJ.47.02.12

This is an open-access article distributed under the terms of a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International Public License, which permits unrestricted use, adaptation, distribution and reproduction in any medium, non-commercially, provided the new creations are licensed under identical terms as the original work and the original work is properly cited.

diagnosing pathological patterns of organs such as thyroid, breast, liver, heart, muscles, and vessels [4].

The great disadvantage of ultrasound is the strong operator dependence. In order to resolve this issue, artificial and machine learning techniques could be developed.

As a subfield of machine learning and artificial intelligence, the deep learning domain is a fascinating and powerful tool for computer vision.

Its accelerated growth is due to the development of powerful hardware, new optimized techniques, software libraries, and large datasets.

The deep learning techniques are very powerful due to the artificial neural networks with many layers designed to extract a hierarchy of discriminant features from raw images.

Deep learning with convolutional neural networks becomes a promising and robust tool in ultrasound imaging classification, detection, and segmentation [5].

Practical ultrasound applications of deep learning have been developed by researchers in the field. In [6], it was developed an architecture of two cascaded CNN models for the assessment of fetal ultrasound images.

Ma et al. [7] proposed a convolutional neural network approach to automatically segment thyroidal nodules from ultrasound images.

Yu et al. [8] developed a CNN architecture with 16 layers to classify the fetal US plane.

In [9] it was proposed a method based on Single Shot MultiBox Detector (SSD), for detecting tumors in breast ultrasound images.

In [10] it was used a pre-trained VGG architecture to classify the anatomic location and plane of abdominal ultrasound images.

In our study, we used the potential of two powerful deep learning pre-trained architectures, Inception v3 and VGG-19, to differentiate the autoimmune disorder in US images.

We designed and re-purposed these architectures using transfer learning with fine-tuning.

Materials and Methods

Method Description

The overall architecture of our methods is illustrated in Figure 1, having the following pipeline:

- The US thyroid nodules images are preprocessed by transformations as resizing and shifting. The result was an augmented dataset.
- The augmented dataset was split into train and test sets.
- We trained our pre-trained CNN models on 100 epochs, using the training dataset.
- The performance of the proposed methods was measured using the test dataset, with the following evaluation metrics: accuracy, AUC-ROC, specificity, sensitivity, and Precision/Recall.

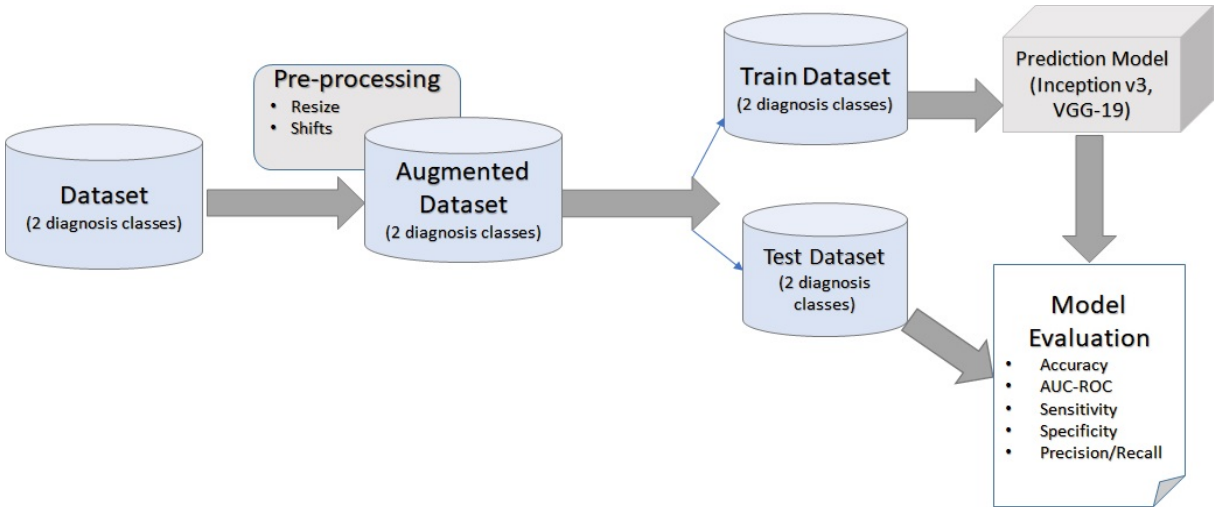

Figure 1. The overall architecture of the proposed method.

Transfer Learning with Fine-Tuning

Transfer learning introduces a different training prototype of using already pre-trained classifiers as a starting point for a new classification problem.

For this purpose, the convolutional neural networks (Inception, VGG, ResNet, etc.) were trained to recognize more than 1,000 classes from large-scale datasets [11-13].

Then we could use the same convolutional neural networks to recognize, for example, medical images.

These deep neural networks were trained on large-scale datasets such as ImageNet and demonstrated to be very performant for transfer learning.

Being trained to learn a large set of representative features, these architectures can be re-purposed for classification problems other than those they were originally trained on.

Design and Implementation of the Proposed Pre-Trained Models

The first method we proposed was based on the pre-trained Inception v3 network.

The Inception module introduced by Szegedy et al. [11] is a complex architecture with many improvements to increase the performance in terms of speed and accuracy. The idea behind the Inception module was of a multi-level feature extractor by computing 1x1, 3x3, and 5x5 convolution operations in the same module of the network. The weights for Inception v3 are of 96MB size, being smaller than VGG architectures [12].

We designed and re-purposed the Inception v3, by transfer learning with fine-tuning, in the following steps:

- removing the original classifier of the network.
- adding at the top of the network one fully connected layer with 256 hidden units and ReLU activation function.
- choosing to train the top 2 inception blocks: we have frozen the first 249 layers and unfrozen the rest.
- adding the last fully connected layer the with softmax activation function for classification.
- choosing RMSprop with a learning rate of 0.001 as optimizer [14].

The second pre-trained method was based on the VGG-19 pre-trained architecture. VGG-19 is an architecture that consists of 19 layers from which, 16 layers are convolution layers, 3 layers are fully connected, 5 layers are maximized pool and one layer is a softmax layer [14].

We designed and re-purposed the VGG-19, by transfer learning with fine-tuning, in the following steps:

- replacing the original classifier of the VGG-19 architecture.
- freezing the convolutional base and then used its outputs to feed the diagnosis classifier.
- adding at the top of the network one fully connected layer of 512 hidden units and ReLU activation function.
- adding the last layer as a softmax dense layer with 2 hidden units used for classification.
- using the Adam with a learning rate of 0.0001 as an optimizer [15].

The input of the two networks was a fixed size of (224 * 224) gray image, which means that the matrix was of shape (224, 224, 3). The images flowed in batches of 32.

Patients and Image Dataset

Our study included 123 patients who were referred for thyroid ultrasound to the Craiova Endocrinology Private Clinic, Pediatric Clinic of County Hospital of Craiova, Endocrinology Clinic of the Municipal Hospital of Craiova.

Eighty-three patients had an autoimmune thyroidal aspect on ultrasound, and 40 had a normal aspect.

Every person included in the study signed a written informed consent form.

Prior to the procedure, the age, gender, background area, weight, current medication, and other conditions were recorded for each patient.

None of the patients included in the study were pregnant or breast-feeding, and none were under treatment with Tyrozol or Euthyrox.

The patients' ages ranged from 17 days to 75 years, with a slight predominance of the female gender.

Thyroid ultrasounds were performed by endocrinologists with a minimum of 20 years' experience using a low-medium-resolution (7.5 MHz) ultrasound device (Siemens SonoAce) equipped with a linear transducer.

The patient remains comfortable throughout the exam, which normally takes only a few minutes unless the lateral neck needs to be evaluated.

The test does not need the patient to discontinue any medication or to be prepared.

The procedure is generally performed with the patient in supine position and the neck hyperextended, but it can also be performed seated.

Multiple laboratory tests were performed to evaluate thyroid function: TSH (thyroid stimulating hormone), free T4 (thyroxine), thyroid peroxidase and thyroglobulin antibodies, blood glucose levels, calcium, liver function, and vitamin D intake.

Additional dosage of Thyrotropin Receptor Antibodies (TRAb) was recommended for patients with hyperthyroidism. Thyroid volume, ultrasound pattern, heterogeneity, nodule presence and size, and the presence of pseudo-nodules and cysts were all observed on ultrasound.

The sonographic appearance of Autoimmune thyroiditis is well recognized, the gland is often diffusely enlarged, and the parenchyma is coarsened, hypoechoic, and often hypervascular.

All images were classified into the autoimmune aspect and normal (Figure 2).

For each patient, 5 thyroid ultrasound images were performed by an endocrinologist, in cross and longitudinal sections for each lobe.

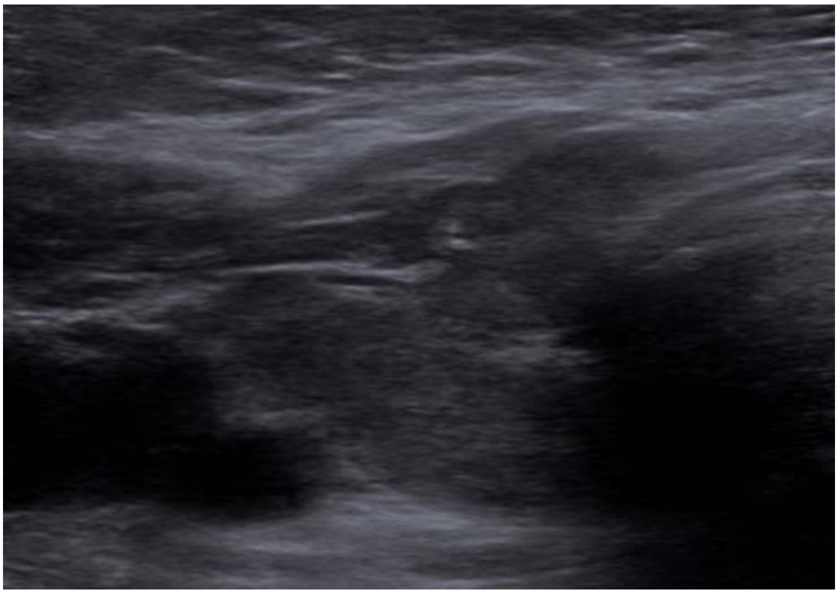

Figure 2. Examples of thyroidal US images collected from patients: (a) autoimmune; (b) normal.

The dataset contains 615 grayscale thyroid images with a final resolution of 224x224 pixels, after removing the patients ’details.

We used 495 images from 99 patients for training and 120 images from 24 patients for testing. The distribution of patients and images among the diagnosis are described in Table1.

Table1. The distribution of images and patients in the training and testing datasets.

Diagnosis	Training		Testing		Total	
	No Patients	No Images	No Patients	No Images	No Patients	No Images
Autoimmune	67	335	16	80	83	415
Normal	32	256	8	40	40	200
Total	99	495	24	120	123	615

Evaluation Metrics

The following metrics were used to evaluate the classification performance: specificity (Sp) (1), sensitivity or recall (Se) (2), the test accuracy (Accuracy) (3), and the area under the ROC curve.

$$Sp = \frac{|true\ negative|}{|true\ negative| + |false\ positive|} \quad (1)$$

$$Se = \frac{|true\ positive|}{|true\ positive| + |false\ negative|} \quad (2)$$

$$Accuracy = \frac{|correctly\ classified\ cases|}{|test\ cases|} \quad (3)$$

We also computed the Precision-Recall curves for each diagnosis class. The precision shows how accurate our model was from predicted positive cases.

The recall or sensitivity calculates how many of the actual positive cases our model captured.

Experimental Results Analysis

Experimental Setup

The thyroid ultrasound images dataset of 615 images was randomly split into the training dataset and test dataset (80% of images for training and 20% of images for testing).

We designed two experiments in order to differentiate the autoimmune from normal US images:

1. Training and assessing the Inception v3 model on the train (563 images) and test datasets (138 images)
2. Training and assessing the VGG-19 model on the train (563 images) and test datasets (138 images)

Results

We compared the diagnosis performance of the two pre-trained models (Inception v3 and VGG-19) in the test dataset.

For the Inception v3 classification model, our results showed an overall diagnosis accuracy of 96.4% and an AUC of 0.96, while for VGG-19 classification model we recorded the overall diagnosis accuracy of 98.6% and an AUC of 0.98.

Better overall performance was achieved when using the VGG-19 architecture as compared with the Inception v3 architecture.

The Inception v3 recorded a maximum sensitivity of 100% in differentiating the autoimmune cases with a specificity of 91.17%.

The VGG-19 model assured a very good balance between precision (95.94%) and specificity (100%) in differentiating the autoimmune patterns in thyroid images.

All metrics computed for our models are summarized in Table 2 and Table 3 and showed accurate results achieved by using the VGG-19 model with transfer learning.

Table 2. Evaluation metrics for classification of thyroid nodules images with Inception v3 model.

Diagnosis	**Accuracy (%)**	**AUC**	**Sensitivity (%)**	**Specificity (%)**
Autoimmune	96.40	0.96	100	91.17
Normal	96.40	0.96	91.17	100
Average	96.4	0.96	95.58	95.58

Table 3. Evaluation metrics for classification of thyroid nodules images with VGG-19 model.

Diagnosis	Accuracy (%)	AUC	Sensitivity (%)	Specificity (%)
Autoimmune	98.80	0.98	95.94	100
Normal	98.40	0.99	100	97.89
Average	98.60	0.98	97.97	98.94

The precision-recall curves for Inception v3 and VGG-19 models could be observed in Figure 3.

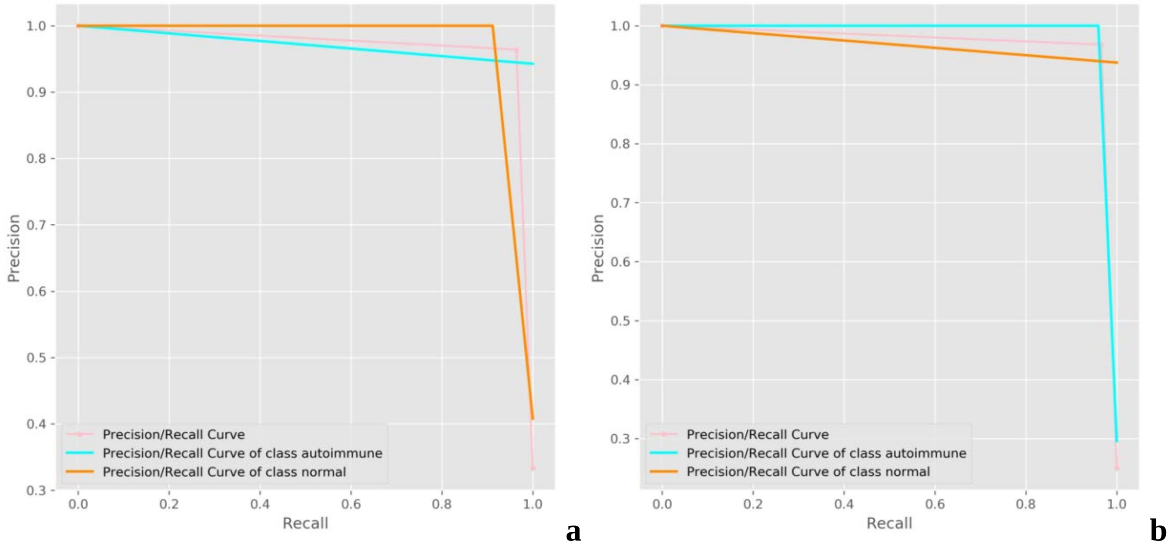

Figure 3. The Precision/Recall curves for each diagnosis class: a) Inception v3; b) VGG-19.

Discussion

Even though the deep learning system used for the analysis of ultrasound had shown promising performance, further improvement is expected.

The current deep learning system cannot accurately accomplish tasks that are impossible for endocrinologists but may also make mistakes that a practicing physician will not [16].

The most reliable method of imaging for the diagnosis of thyroid disorders is Ultrasonography (US).

Despite its significance, numerous studies have revealed that US diagnostic quality is variable and is influenced by the operator's experience [17-21].

As a result, a new CAD model based on artificial intelligence was expected to boost US diagnostic efficiency while reducing interobserver variability through the use of a semi-automated workflow [22-27].

Concerning the CAD system's diagnostic performance, multiple studies have found comparable diagnostic performance between the CAD system and experienced radiologists [23,24].

By analyzing the results, our proposed methods based on deep learning with transfer learning and fine-tuning had significant results in the differentiation of thyroid images.

Our transfer learning VGG-19 model can distinguish autoimmune aspects from normal cases with a great overall accuracy of 98.6%, while the Inception v3 model recorded an overall accuracy of 96.4%.

Computer-aided diagnosis (CAD) systems have been recently applied in the differential diagnosis of thyroid nodules and disorders and show a great potential to reduce human dependence and improve the classification performance of US thyroid images.

Our pre-trained CNN methods used for differentiating autoimmune thyroid aspect from normal aspect had some limitations:

- the testing dataset was known and there was a risk of overfitting and weak generalization;
- the dataset was imbalanced and had fewer images for thyroid normal aspect;
- the ultrasound images did not have the best resolution.

Conclusion

In this study, we focused on finding a way to differentiate the autoimmune thyroid aspect from the normal thyroidal aspect.

For this purpose, we developed a novel DL algorithm that is superior to the traditional ML methods as it does not require an initial description of visual features of medical images.

The experimental results indicated that the proposed VGG-19 architecture outperforms the Inception v3 network, with a precision of 95.94% and specificity (100%) in differentiating the autoimmune aspects from the normal in thyroid images.

In clinical practice, the methods based on deep learning algorithms could become effective diagnosis tools in assisting endocrinologists.

Abbreviations

AITD - Autoimmune thyroid disease
HT - Hashimoto's thyroiditis
GD - Graves Disease
US - Ultrasound
CAD - Computer-Aided Diagnosis
ML - Machine Learning
DL - Deep Learning
CNN - Convolutional Neural Network
AUC - Area Under the Curve

Institutional Review Board Statement

The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of University of Medicine and Pharmacy of Craiova (approval no. 6/20.01.2021).

Informed Consent Statement

Informed consent was obtained from all subjects involved in the study. Written informed consent has been obtained from the patient(s) to publish this paper.

Acknowledgments

The research has received funding from Norwegian Financial Mechanism 2014-2021 under the project RO-NO-2019-0138, 19/2020 “Improving Cancer Diagnostics in Flexible Endoscopy using Artificial Intelligence and Medical Robotics” IDEAR, Contract No. 19/2020.

Corina Maria Vasile and Anca Loredana Udriştoiu, share first authorship, their contribution on study design and manuscript preparation was equal.

Conflicts of Interests

The authors declare no conflict of interest.

References

1. McLeod DS, Cooper DS. The incidence and prevalence of thyroid autoimmunity. Endocrine, 2012, 42(2):252-265.
2. Kim D. The Role of Vitamin D in Thyroid Diseases. Int J Mol Sci, 2017, 18(9):1949.
3. Ilka Y, de Cássio SO, Chammas MC, Cerri GG. Ultrasound findings in thyroiditis. Radiol Bras, 2007, 40(2):75-79.
4. Kim M, Yun J, Cho Y, Shin K, Jang R, Bae HJ, Kim N. Deep Learning in Medical Imaging. Neurospine, 2019, 16(4):657-668.
5. Liu S, Wang Y, Yang X, Lei B, Liu L, Li SX, Ni D, Wang T. Deep Learning in Medical Ultrasound Analysis: A Review. Engineering, 2019, 5(2):261-275.
6. Wu L, Cheng J-Z, Li S, Lei B, Wang T, Ni D. FUIQA: fetal ultrasound image quality assessment with deep convolutional networks. IEEE Trans Cybern, 2017, 47(5):1336-1349.
7. Ma J, Wu F, Jiang T, Zhao Q, Kong D. Ultrasound image-based thyroid nodule automatic segmentation using convolutional neural networks. Int J CARS, 2017,12(11):1895–1910.
8. Meng D, Zhang L, Cao G, Cao W, Zhang G, Hu B. Liver fibrosis classification based on transfer learning and FCNet for ultrasound images. IEEE Access, 2017, 5:5804–5810.
9. Cao Z, Duan L, Yang G, Yue T, Chen Q, Fu H, Xu Y. Breast tumor detection in ultrasound images using deep learning. In: Proceedings of International Workshop on Patch-based Techniques in Medical Imaging. Springer, 2017, Cham, 121-128.
10. Cheng PM, Malhi HS. Transfer learning with convolutional neural networks for classification of abdominal ultrasound images. J Digit Imaging, 2017, 30(2):234-243.
11. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2016, Las Vegas, 2818-2826.
12. Simonyan, K, Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition, 2015 [online]. Available at: https://arxiv.org/abs/1409.1556. [Accessed 01/03/2020].
13. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, 770–778.
14. Keras, Deep Learning for Humans [online]. Available at: https://github.com/keras-team/keras. [Accessed 1/06/2020].
15. Kingma DP, Ba J Adam. A method for stochastic optimization, 2014 [online]. Available online at: https://arxiv.org/abs/1412.6980. [Accessed 1/05/2020].
16. Launchbury J. A DARPA perspective on artificial intelligence [online]. Available at www.darpa.mil/attachments/AIFull.pdf. [Accessed 03/11/2020].
17. Choi SH, Kim EK, Kwak JY, KimMJ, Son EJ. Interobserver and intraobserver variations in ultrasound assessment of thyroid nodules. Thyroid, 2010, 20:167–172.
18. Kim HG, Kwak JY, Kim EK, Choi SH, Moon HJ. Man to man training: can it help improve the diagnostic performances and interobserver variabilities of thyroid ultrasonography in residents? Eur J Radiol, 2012, 81:e352–e356.
19. Park CS, Kim SH, Jung SL, Kang BJ, Kim JY, Choi JJ, Sung MS, Yim HW, Jeong SH. Observer variability in the sonographic evaluation of thyroid nodules. J Clin Ultrasound, 2010, 38(6):287-293.
20. Park SH, Kim SJ, Kim EK, Kim MJ, Son EJ, Kwak JY. Interobserver agreement in assessing the sonographic and elastographic features of malignant thyroid nodules. Am J Roentgenol, 2009, 193:W416–W423.
21. Kim SH, Park CS, Jung SL, Kang BJ, Kim JY, Choi JJ, Kim YI, Oh JK, Oh JS, Kim H, Jeong SH. Observer variability and the performance between faculties and residents: US criteria for benign and malignant thyroid nodules. Korean J Radiol, 2010, 11(2):149.
22. Acharya UR, Sree SV, Krishnan MM, Molinari F, ZieleŸnik W, Bardales RH, Witkowska A, Suri JS. Computer-aided diagnostic system for detection of Hashimoto thyroiditis on ultrasound images from a Polish population. J Ultras Med, 2014, 33(2):245-253.
23. Chang Y, Paul AK, Kim N, Baek JH, Choi YJ, Ha EJ, Lee KD, Lee HS, Shin D, Kim N. Computer-aided diagnosis for classifying benign versus malignant thyroid nodules based on ultrasound images: a comparison with radiologist-based assessments. Med Phys, 2016 ,43(1):554-567.
24. Choi YJ, Baek JH, Park HS, Shim WH, Kim TY, Shong YK, Lee JH. A computer-aided diagnosis system using artificial intelligence for the diagnosis and characterization of thyroid nodules on ultrasound: initial clinical assessment. Thyroid, 2017, 27(4):546-552.
25. Li LN, Ouyang JH, Chen HL, Liu DY. A computer aided diagnosis system for thyroid disease using extreme learning machine. J Med Syst, 2012, 36(5):3327–3337.
26. Lim KJ, Choi CS, Yoon DY, Chang SK, Kim KK, Han H, Kim SS, Lee J, Jeon YH. Computer-aided diagnosis for the differentiation of malignant from benign thyroid nodules on ultrasonography. Acad Radiol, 2008, 15(7):853-858.
27. Chen KY, Chen CN, Wu MH, Ho MC, Tai HC, Kuo WH, Huang WC, Wang YH, Chen A, Chang KJ. Computerized quantification of ultrasonic heterogeneity in thyroid nodules. Ultrasound Med Biol, 2014, 40(11):2581-2589.

Corresponding Authors: Alice Elena Ghenea, Department of Bacteriology-Virology-Parasitology, University of Medicine and Pharmacy of Craiova, Romania, e-mail: gaman_alice@yahoo.com

Vlad Padureanu, Department of Internal Medicine, University of Medicine and Pharmacy of Craiova, Romania, e-mail: vladpadureanu@yahoo.com