

Deep learning in the design and application of thoracic medical image-assisted diagnosis algorithm

Yineng Xiao*

^a The Global Intellectual Property Institute, Nanjing University, Suzhou, 215163, China;

^b Advanced Institute of Information Technology, Peking University, Hangzhou, 311200, China.

ABSTRACT

In recent years, with the development of deep learning (DL) technology and the continuous improvement of algorithms, DL-assisted diagnosis systems based on medical images have rapidly developed. Compared with traditional image processing methods, DL trains models by combining a large amount of relevant data (e.g. clinical data, imaging data, etc.) and then uses the models to predict disease-related information. Compared with traditional medical image processing algorithms, DL has better performance in recognition, segmentation and classification of medical images. In this paper, an AI-aided diagnosis algorithm is developed for lung cancer, a malignant tumour disease, based on clinical chest CT data and imaging data. The algorithm uses chest CT images as the object of study, and classifies and evaluates patients based on various factors such as their age, gender, tumour volume and location, as well as their knowledge of the disease. A pre-trained model was first used to establish an algorithm for lung cancer tumour segmentation and recognition. Convolutional NNs are then applied to learn solutions to the feature extraction and classification problems. Finally, the results obtained are used as model output and the system performance is evaluated to complete the diagnostic process for lung cancer as a class of malignant tumour disease.

Keywords: Deep learning, neural networks, thoracic medicine, image assist, diagnostic algorithms

1. INTRODUCTION

In chest imaging diagnosis, chest CT images are the most commonly used medical imaging data, which can display and locate important information such as the site, location and extent of lesions¹⁻². In a chest medical examination, doctors look at a patient's chest CT image to determine his or her condition. For patients with diseases such as lung cancer, doctors usually need to observe the growth and extent of the tumour; for lung cancer patients, they also need to observe lung function and some other imaging indicators to assess the condition. In CT images, the asymmetry or irregularity of lung lesions makes the location of lesions vary from region to region; and the shape of lung lesions also changes with the flow of gas in the lungs, so there is information overlap between different regions. In the field of computer-aided diagnosis, DL methods are mainly used for computer vision tasks in order to perform image classification and recognition. DL algorithms have also been widely used in the field of medical imaging, such as the combination of DL techniques and medical image processing algorithms for tumour localisation, tumour benignity and malignancy determination, and some other disease diagnosis in the medical field³⁻⁴.

*xiaoyineng@pku.edu.cn

In a related study, Ilyas et al. developed a new patient-specific anatomical background and shape a priori or PACS-aware 3D recurrent registration-segmentation network for segmentation of longitudinal thoracic CBCT⁵. The segmentation and registration networks were trained simultaneously in an end-to-end framework and implemented by convolutional long and short-term memory models. The registration network is trained in an unsupervised manner using planned CT (pCT) and CBCT image pairs and produces a progressively deformed image sequence. The segmentation network was optimised by combining progressively deformed pCT (anatomical background) and pCT delineations (shape background) with CBCT images in a one-time setup. John et al. proposed a new DL framework to explore discriminative information in the lung and heart regions⁶. A feature extractor equipped with a multi-scale attention module was designed to learn global attention maps from global images. Comprehensive experiments show that our approach achieves superior performance compared to state-of-the-art methods. The proposed new deep framework for multi-label classification of chest diseases in chest X-ray images. The proposed network aims to efficiently exploit the pathological regions containing the main clues from chest radiography. The proposed network has been used for clinical screening to assist radiologists. Chest X-rays represent a large proportion of radiological examinations and there is value in exploring additional ways to improve performance.

Lung cancer is one of the most common malignant tumours in China, accounting for about 40% of all malignant tumours. Due to the lack of obvious symptoms in the early stages of lung cancer and the small size of the tumour, many patients are already in the middle to late stages when their disease is detected. The use of X-rays in CT examinations cannot determine the extent of the lesion and whether further treatment is needed. Based on this, this paper develops and clinically validates a system based on DL algorithms to assist in the detection of lung cancer. In this paper, a data-driven artificial intelligence model is constructed using a convolutional NN algorithm to assist doctors in the diagnosis of lung cancer and other malignant diseases. The results are then used as input to the model to predict the patient's condition.

2. DESIGN STUDIES

2.1 Deep Neural Network (DNN) application problems

The development of big data has driven the rise of DL, however, unlike natural image datasets, medical data is difficult to collect in a standardised way, and data annotation requires a high degree of specialisation and is expensive, making it very difficult to obtain large-scale annotated data. In addition, feature learning in deep networks is closely related to the class distribution of the dataset, and the very limited and unbalanced medical data is prone to problems such as biased feature selection and overfitting. Therefore, the application of DNNs to fine-grained chest X-ray aided diagnosis of high standard still faces two critical problems as follows⁷⁻⁸.

2.1.1 Inadequate sample of labelled data

Firstly, the problem of difficulty in fitting features to the network due to insufficient samples of labeled data. Although ChestX-Ray14 is a relatively large medical dataset with a data volume of 110,000, compared to the natural image dataset Imagenet with a data volume of 10 million, networks trained on medical image datasets still suffer from poor generalization and overfitting due to insufficient sample size.

If the target dataset is not sufficiently supportive, fine-tuning the full layer may result in over-fitting the network; or fine-tuning the fully connected layer only due to the low similarity between the source and target domain data may prevent the network from extracting the feature semantic information of the chest lesion accurately inductively and with poor expressiveness.

2.1.2 Uneven distribution of case data

Not only is the number of normal X-ray images generally higher than those containing lesions, but due to factors such as the complex diversity of disease pathogenesis, there may be significant biases in the distribution of samples for certain diseases, with large disparities in the number of samples for different categories of disease. The problem of data imbalance is currently addressed at two main levels⁹⁻¹⁰.

(1) Data: There are two ways to change the distribution of data through data pre-processing: one is to oversample a few (abnormal) classes and the other is to undersample most (normal) classes, but most oversampling methods do not introduce additional data into the model, so increasing the amount of data from a few classes through oversampling is

difficult to increase the information contained in the data and may easily cause model overfitting. Under-sampling and discarding some of the data may result in the loss of information on key focal area features and make the model more biased. In addition, medical imaging data is inherently scarce and discarding data is impractical and extravagant.

(2) Algorithmic aspects: the learning direction of the network can be changed by the loss function. For DNNs, too many negative samples (disease-free samples) that are simple and easy to divide will change the direction of network gradient update during the learning process, making the network biased towards the learning of normal samples, generating the problem of feature learning bias and reduced feature extraction ability of critical lesion regions.

2.2 Thoracic medical imaging assisted diagnosis system

The thoracic medical imaging assisted diagnostic system in this paper consists of the following main areas.

(1) User management functions: There are two types of users in the system, including general users and system administrator users. The general user is the front-line clinical radiologist: the system administrator is used to manage user information within the system, update the system model and maintenance, etc.

(2) Auxiliary diagnostic functions and business extension functions: The auxiliary diagnostic functions of medical images are the core functions within the application system of this chapter, completing the inference of the models implemented in the previous two chapters for thoracic medical images. At the same time, when designing the system, it is also necessary to fully consider the scalability of the business functions of the system, and all services are implemented in the form of components, so that new diagnostic services can be extended around medical images of the thorax or other parts of the body.

(3) Persistence function of patient diagnostic data: Only by constructing a persistence mechanism in some form for structured personal patient information as well as unstructured medical imaging data can we achieve continuous observation of patient conditions and facilitate physicians to view historical patient condition information and give the best diagnostic advice.

(4) Scalability of image data access methods: In most cases, multiple tomographic scanner devices exist in the actual hospital environment. Therefore, not only do we need to use a reasonably designed image file storage system to cope with the massive amount of medical image data, but we also need to consider the difference in image formats of different data sources, such as MHD format and DICOM format images¹¹⁻¹².

2.3 Cross-Entropy loss function

Cross entropy is often used as a loss function in DL. The loss function reflects the distance between the output value y' and the label value y . A smaller gap indicates a better fit.

In the logistic regression task, let the input be x and the output be y' , the linear regression model can be abbreviated as

$$y' = w^T x + b \quad (1)$$

and define a non-linear mapping function from the input space to the output space as in equation (2).

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (2)$$

For a 2 classification problem, where the label value y takes the value 0 or 1, the posterior probability function at $y = 1$ can first be defined as

$$P(y = 1 | x) = y' = \sigma(w^T x + b) \quad (3)$$

The posterior probability at $y = 0$ is then defined as

$$P(y = 0 | x) = 1 - y' = 1 - \sigma(w^T x + b) \quad (4)$$

In summary, the posterior probability $P(y|x)$ can be defined as follows.

$$P(y | x) = y'^y (1 - y')^{(1-y)} \quad (5)$$

According to the maximum likelihood estimate, if all samples satisfy independent identical distribution, then a set of parameters can be identified to maximise the value of equation (5). Since the logarithmic function is monotonically increasing, then maximising the value of $P(y|x)$ can be equated to maximising the value of $\log(P(y|x))$, so taking the logarithm of equation (5) yields

$$\log(P(y | x)) = y \log(y') + (1 - y) \log(1 - y') \quad (6)$$

The objective of logistic regression is to maximise this function, which adequately reflects the model performance error. The loss function can therefore be made to be the opposite of the above function, and for m samples the loss function is the cross-entropy function.

$$L(w, b) = -\frac{1}{m} \sum_{i=1}^m y \log(y') - (1 - y) \log(1 - y') \quad (7)$$

3. EXPERIMENTAL RESEARCH

3.1 Experimental environment and parameter settings

The PyTorch framework is a commonly used DL framework in which one can design one's own network model and automatically solve the gradient by back propagation within the framework. The reduction of the U-Net network framework and the training of the DB-U-Net network are based on the PyTorch framework. A pre-configured runtime environment and library files are used to run the lung CT image pre-processing algorithm and the graph cut algorithm.

The basic configuration of the experimental environment in this paper is an Intel i5-8400 processor, two NVIDIA GeForce 1070Ti graphics cards (8GB of video memory), 32G of RAM, and Ubuntu 16.04 Server Edition. The GPU parameters, such as computational power and speed, will affect the network training of the experiments, and the use of GPU performance will reduce the training time of the split network in this paper.

The experimental parameters are shown in Table 1. Because the lung CT images used in this paper are relatively large, the number of samples selected for the same training session will have the problem of overflowing video memory, which will affect the calculation results and the calculation speed. The effectiveness of the training depends heavily on the learning rate, which is set to 0.01, 0.001, 0.0001 and 0.00001, respectively.

Table 1. Experimental parameter settings.

Parameter Name	Parameter values
Learning rate	0.0001
Batch size	16
Number of iterations	50
Momentum	0.9
Weight decay	0.0001

Stochastic gradient descent is used for training the model in this chapter to solve for the minima, updating the model weights in each iteration until it gradually converges to the minima. In a single iteration, the weights of the model are updated using a batch size number of samples each time. The advantages of using stochastic gradient descent over other optimisation algorithms are as follows: (1) it is possible to train the network model without loading all the data into memory at once, taking up less memory and video memory; (2) stochastic gradient descent algorithms train the network faster; (3) within reason, increasing the batch size can make the direction of gradient descent more accurate and faster, as smaller batches can cause the model to fall into other locally optimal solutions. When training the NN, the batch size was set to 16.

3.2 General flow of the experiment

In this paper, a feature interpretability-guided multi-scale integrated convolutional NN is proposed to achieve abstract feature reuse with understanding of cnn depth features, and to accurately and automatically perform the task of classifying true and false recurrence of glioma in DTI images. Firstly, three single classification models for true and false recurrence of glioma were trained separately using the classical cnn framework as the base model. All feature maps are then visualised layer by layer, and the heatmap mapping approach is used to visually capture the respective "focus" of the different layers of the single classification models. For example, some layers focus more on edge information, while others highlight differences between pixels, etc. Based on this, the three monclassification models are empirically observed by the imaging practitioner to locate the most relevant layer for each glioma lesion area, visually completing the feature selection process. Finally, the selected multi-scale features from different networks were fused to construct an integrated model of true and false recurrence of glioma. The experimental flow chart is shown in Figure 1.

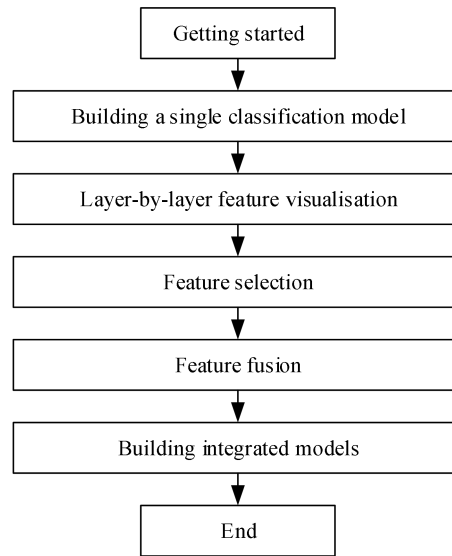


Figure 1. Experimental flow chart.

3.3 Performance evaluation indicators

In order to verify the accuracy of the algorithm based on DL and feature mixing, the performance of the algorithm was measured based on the categories to which the sample labels belonged compared with the actual predicted categories of the algorithm using the evaluation criteria shown in Table 2 to overcome the singularity and one-sidedness of the algorithm performance evaluation in the field of medical aid diagnosis in terms of accuracy alone.

Table 2. Evaluation criteria for detection of lung nodules.

Test	Results	Gold Standard
	Positive	Negative
Positive	TP	FP
Negative	FN	TN

The meanings of the indicators in Table 2 are as follows.

(1) TP (true positive) for a true positive i.e. positive sample (lung nodule lesion area) is correctly determined by the algorithm as a positive sample (lung nodule lesion area).

(2) FP (false positive) is a false positive, i.e. a negative sample (non-lung nodule area) is incorrectly determined by the algorithm to be a positive sample (lung nodule lesion area)

(3) TN (true negative) is true negative, i.e. a negative sample (non-lung nodule region) is correctly determined by the algorithm to be a negative sample (non-lung nodule region).

(4) FN (false negative) is a false negative, i.e. a positive sample (lung nodule lesion area) is incorrectly determined by the algorithm as a (negative sample (non-lung nodule area).

In the lung nodule detection algorithm, the ultimate aim is to extract all the lung nodules in the lung CT image sequence, but since interference from other tissues within the lung often causes a degree of false positives, the experiment uses the accuracy (ACC), sensitivity, and specificity (Spe) defined by the four criteria mentioned above for a comprehensive measure of the overall discriminatory ability of the algorithm.

The accuracy is used to measure the overall classification capability of the algorithm, and the higher the accuracy, the better the overall classification of the algorithm, as defined in equation (8).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} * 100\% \quad (8)$$

Sensitivity, also known as true positive rate and recall, is used to measure the ability of the algorithm to discriminate between regions of lung nodules, i.e. the ratio of the number of correctly predicted samples among all nodule samples, with higher sensitivity representing a lower rate of missed detections by the algorithm, as defined by the formula in equation (9).

$$Sensitivity = \frac{TP}{TP + FN} * 100\% \quad (9)$$

Specificity is used to measure the ability of the algorithm to discriminate between non-nodular regions, i.e. the ratio of the number of correctly predicted samples out of all non-nodular samples, with a higher specificity representing a better determination of non-nodular regions, as defined by the formula in equation (10).

$$Specificity = \frac{TN}{TN + FP} * 100\% \quad (10)$$

In this paper, we use the FROC (Free Response Receiver Operating Characteristics) criterion to measure the detection effectiveness of the algorithm by calculating the cpm (Competitive Performance Index) value, where cpm is the average number of false positives per set of CT images (FPs per scan) i.e. the average detection rate at horizontal coordinates of 1/8, 1/4, 1/2, 1, 2, 4, 8. The following experiments use cpm values as an assessment criterion.

4. EXPERIMENTAL ANALYSIS

4.1 Experimental data sources

The dataset was derived from lung CT image segmentation, with 7,500 CT images, 6,000 for training and validation, which contains 4,800 training set data and 1,200 validation set data. The training and validation datasets were all collected from the National Centre for Biological Information and contained a dataset of 2000 normal CT line images, 2000 CT line images of viral pneumonia and 2000 CT line images of positive new coronary pneumonia as experimental subjects. Details of the dataset are shown in Table 3 for a total of 6000 CT images of lung data, 80% of the data is the training set data and 20% of the data is the validation set data.

Table 3. Lung CT image dataset

Classification	Normal CT images	CT image of viral pneumonia	Positive CT image of New Coronary Pneumonia
Total	2500	2500	2500

Training set	1600	1600	1600
Validation set	400	400	400
Test set	500	500	500

4.2 Analysis of experimental results

To validate the effectiveness of the proposed DL and feature mixing based lung nodule detection model, CT image sequences were sent to the network for detection, and patient image data from the test set were selected for the following self-comparison experiments and statistical quantification of the prediction results by calculating cpm values using guidelines to measure the performance of the algorithm: (1) Lung nodules were detected using the traditional R-fcn network and the results were recorded as modell. Table 4 shows the difference in performance between the three network structures on the LUNA16 dataset for Figure 2 shows the FROC curves of the three models for the improved process.

Table 4. Algorithm improvement process detection performance comparison

	Model 1	Model 2	Model 3
0.125	0.64	0.72	0.74
0.25	0.70	0.80	0.83
0.5	0.79	0.83	0.89
1	0.83	0.86	0.90
2	0.89	0.91	0.92
4	0.91	0.92	0.94
8	0.92	0.94	0.97
CPM	0.81	0.85	0.88

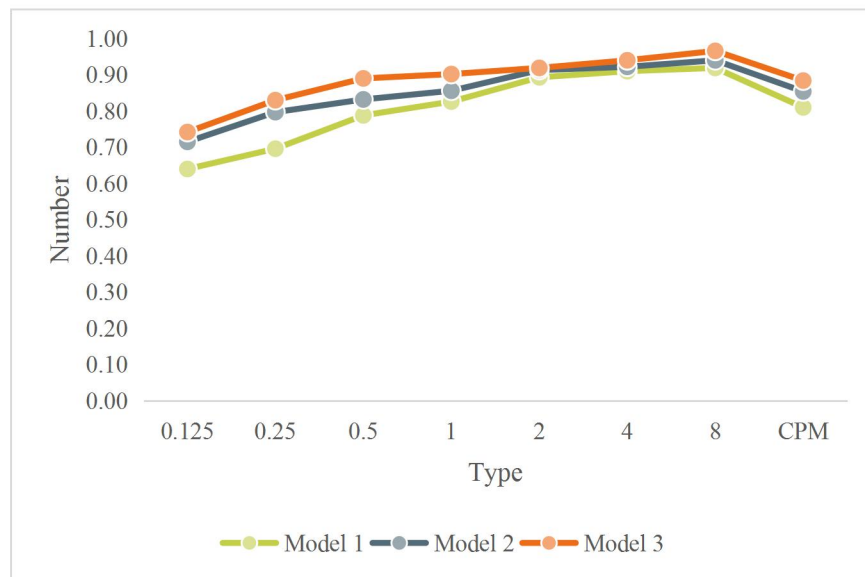


Figure 2. Comparative analysis of the detection performance of the algorithm improvement process

The process of the change in the average detection rate of the algorithm during the improvement can be seen by observing Table 4 with Figure 2. By upgrading the ResNet to denseNet with fpn for feature extraction, the algorithm detection sensitivity was improved and the recognition performance of the network was enhanced. Then according to the

results of K-meaning is clustering change the settings in the rpn network with and correlation in the network, and introduce the feature pyramid structure, training using Focalose loss function instead of the traditional cross-entropy cost function to further improve the recognition performance of the network, have a higher detection accuracy for small-scale nodules, so that the network has better results. The above experiments show that the detection sensitivity of the algorithm is improving in the process of improvement, and the average detection rate can reach 96.6% when the average number of false positives in each group of CT images is 8. The recognition performance of the proposed algorithm is better than the original method in the detection of lung nodules.

In summary, the following conclusions can be drawn from the above experiments: even though lung nodules have a series of complex characteristics such as varying sizes, complex shapes and uncertain locations, the proposed DL and feature mixing based lung nodule detection algorithm still shows good detection results for nodules of different scales and good discrimination ability in the face of complex lesion areas. The network not only has a high prediction probability for larger diameter lung nodules to reach a correct judgment, but also for smaller diameter microscopic lung nodules to reach a correct prediction conclusion.

Table 5 shows the performance of the algorithm in this paper compared with the FasterRCNN, the multi-view convolutional NN model and the multi-view deep belief network for lung nodule detection. The FROC curves for the four models are shown in Figure 3.

Table 5. Comparison of lung nodule detection performance by model

	Faster RCNN	Multi-view CNN	Multi-view DBN	Algorithms in this paper
0.125	0.73	0.69	0.73	0.74
0.25	0.74	0.71	0.80	0.83
0.5	0.76	0.81	0.80	0.89
1	0.80	0.86	0.85	0.90
2	0.82	0.90	0.89	0.92
4	0.83	0.91	0.91	0.94
8	0.83	0.92	0.95	0.97
CPM	0.79	0.84	0.85	0.88

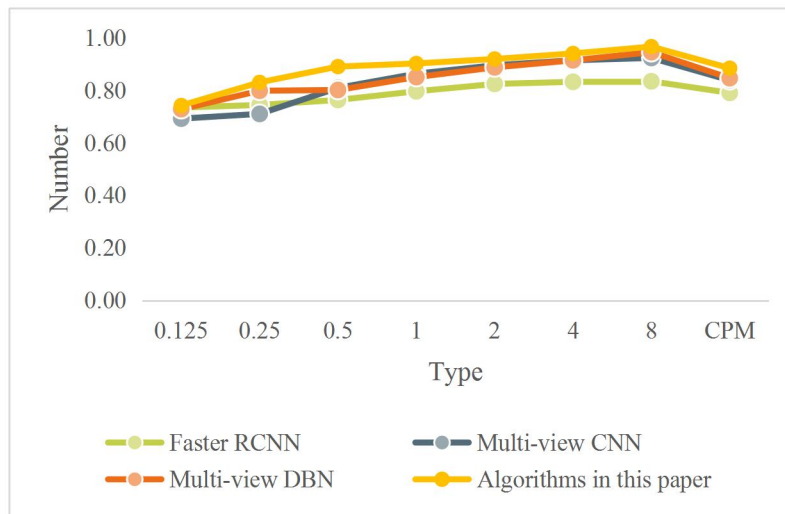


Figure 3. Comparative analysis of lung nodule detection performance by model

Analysis of Table 5 and Figure 3 shows that the proposed algorithm performs well in the task of lung nodule detection, and its overall sensitivity is better than other algorithms, with a sensitivity of 96.6% at an average of 8 false positives per group of CT images. Through the above analysis, the improved algorithm outperforms the traditional algorithm in terms of feature extraction ability and discrimination ability, which can effectively lay the foundation for the subsequent diagnosis of the imaging physician and achieve the purpose of assisting the physician in the diagnosis.

5. CONCLUSIONS

In this paper, a DL algorithm is designed to segment and classify tumours from lung CT images, and the results are tested to show that the algorithm can effectively identify and classify lung tumours and determine their benign and malignant degrees. At present, DL has been widely used in the fields of image recognition, medical image analysis and medical artificial intelligence. The lung CT image segmentation and lung disease diagnosis method studied in this paper is a new artificial intelligence technique based on the application of computer-aided diagnosis (CAA) technology to medical imaging diagnosis, which can classify and diagnose lung tumours more accurately and with more guidance by fusing medical imaging data and clinical experience. The DL algorithm will achieve a more comprehensive and detailed presentation of clinical information than traditional image processing algorithms.

REFERENCES

- [1] Patel, V., Li, C. H., Rye, V., Liu, C. S. J., Lerner, A., Acharya, J., Rajamohan, A. G., A Comparison of WebRTC and Conventional Videoconferencing for Synchronized Remote Medical Image Presentation. *J. Digit. Imaging* 35(1): 68-76 (2022).
- [2] Soomro, T. A., Zheng, L., Afifi, A. J., Ali, A., Yin, M., Gao, J., Artificial intelligence (AI) for medical imaging to combat coronavirus disease (COVID-19): a detailed review with direction for future research. *Artif. Intell. Rev.* 55(2): 1409-1439 (2022).
- [3] Ghasemi, M., Kelarestaghi, M., Eshghi, F., Sharifi, A., D3FC: deep feature-extractor discriminative dictionary-learning fuzzy classifier for medical imaging. *Appl. Intell.* 52(7): 7201-7217 (2022).
- [4] Avola, D., Cinque, L., Fagioli, A., Foresti, G., Mecca, A., Ultrasound Medical Imaging Techniques: A Survey. *ACM Comput. Surv.* 54(3): 67:1-67:38 (2022).
- [5] Sirazitdinov, I., Schulz, H., Saalbach, A., Renisch, S., Dylov, D. V., Tubular shape aware data generation for segmentation in medical imaging. *Int. J. Comput. Assist. Radiol. Surg.* 17(6): 1091-1099 (2022).
- [6] Valen, J., Balki, I., Mendez, M., Qu, W., Levman, J., Bilbily, A., Tyrrell, P. N., Quantifying uncertainty in machine learning classifiers for medical imaging. *Int. J. Comput. Assist. Radiol. Surg.* 17(4): 711-718 (2022).
- [7] Ravi, P., Chepelev, L. L., Stichweh, G. V., Jones, B. S., Rybicki, F. J., Medical 3D Printing Dimensional Accuracy for Multi-pathological Anatomical Models 3D Printed Using Material Extrusion. *J. Digit. Imaging* 35(3): 613-622 (2022).
- [8] Kidav, J., Pillai, P. M., Deepak, V., Sreejeesh S. G., Design of a 128-channel transceiver hardware for medical ultrasound imaging systems. *IET Circuits Devices Syst.* 16(1): 92-104 (2022).
- [9] Benazzouz, M., Benomar, M. L., Moualek, Y., Modified U-Net for cytological medical image segmentation. *Int. J. Imaging Syst. Technol.* 32(5): 1761-1773 (2022).
- [10] Kollem, S., Ramalinga Reddy, K., Srinivasa Rao, D., Rajendra Prasad, C., Malathy, V., Ajayan, J., Muchahary, D., Image denoising for magnetic resonance imaging medical images using improved generalized cross-validation based on the diffusivity function. *Int. J. Imaging Syst. Technol.* 32(4): 1263-1285 (2022).
- [11] Kumar, A., Goodrum, H., Kim, A., Stender, C., Roberts, K., Bernstam, E. V., Closing the loop: automatically identifying abnormal imaging results in scanned documents. *J. Am. Medical Informatics Assoc.* 29(5): 831-840 (2022).
- [12] Valtchinov, V. I., Murphy, S. N., Laeson, R., Ikononov, N., Zhai, B. K., Andriole, K., et al., Analytics to monitor the local impact of the Protecting Access to Medicare Act's imaging clinical decision support requirements. *J. Am. Medical Informatics Assoc.* 29(11): 1870-1878 (2022).