UDC 004.8

IMPACT OF PREPROCESSING AND COMPARISON OF NEURAL NETWORK ENSEMBLE METHODS FOR SEGMENTATION OF THE THORACIC SPINE IN X-RAY IMAGES

Koniukhov V. D. – Postgraduate student, A. Pidhornyi Institute of Power Machines and Systems of NAS of Ukraine, Kharkiv, Ukraine.

Morgun O. M. – PhD, Director of the"Laboratory of X-ray Medical Equipment" LTD, Kharkiv, Ukraine.

Nemchenko K. E. – Dr. Sc., Head of the Department, V. N. Karazin Kharkiv National University, Kharkiv, Ukraine.

ABSTRACT

Context. Automatic segmentation of medical images plays an important role in the process of automating the detection of various diseases in the spine and the use of radiography is the most accessible means of predicting diseases. Over the years many studies have been conducted on the topic of image segmentation. One of the many methods for improving image segmentation is the use of neural network ensembles.

Objective. The aims of this study were to investigate the impact of preprocessing and compare the main methods of neural network ensembles and their effect on the segmentation of the thoracic region, in this study the area was considered which consists of the vertebrae: Th8, Th9, Th10, Th11.

Method. To begin with, the influence of preprocessing of X-ray images was considered, which included the following methods: histogram equalization for contrast enhancement, contrast-limited adaptive histogram equalization, logarithmic transform method, median filter, Gaussian filter, and bilateral filter. To study the influence of neural network ensemble on segmentation quality, several methods were used. Averaging method – a simple half-averaging method. Weighted averaging method – an improved version of the averaging method which uses weights for each network, the higher the network weight, the greater its influence on averaging. Method of cumulative averaging – a modified averaging method in which each ensemble receives an averaged image, after which all the results of the ensembles are averaged. Bagging – method of averaging networks trained on different data, *n* networks are used, the training sample is divided into *n* parts, and each neural network is trained on its own subset of data, as a result, the averaging method is used for predictions. Averaging method for a large number of networks – in this method, 100 neural networks were trained, after which the averaging method was used. Method of averaging mask shapes – this method uses a distance transform to average multiple masks into one shape average.

Results. It was investigated that the use of different methods of image preprocessing does not guarantee an improvement in the quality of segmentation of the spine region on X-ray images, but even on the contrary worsens the quality of segmentation. Different methods of combining predictions of neural network ensembles were considered, which made it possible to find out the pros and cons of specific methods for the task of segmentation of X-ray images.

Conclusions. The experiments conducted allowed us to conclude that the use of any preprocessing methods should not be used for segmentation of X-ray images. Also, due to a large number of architectures and methods for combining predictions, the behavior of ensemble methods was studied, which will allow us to further determine the necessary approach for segmentation of X-ray images. Further study of the weighted averaging method and the mask shape averaging method will make it possible to improve the obtained result and achieve even greater success in segmentation.

KEYWORDS: machine learning; image recognition; neural network; image segmentation, computer vision.

ABBREVIATIONS

BF is a bilateral filter;

CLAHE is a Contrast-Limited Adaptive Histogram Equalization;

DICE is a Dice-Sørensen coefficient;

GF is a Gaussian filter;

HE is a histogram equalization;

IoU is an Intersection over Union;

LT is a logarithmic transformation;

MF is a median filter.

NOMENCLATURE

A is a total number of networks in an ensemble;

$acc(S,P)$ is a method for determining the similarity of two images;

B_M is a block width;

B_N is a block height;

c is a constant that scales the value after a logarithmic transformation;

$dt()$ is a function that calculates for each pixel the distance to the nearest zero pixel;

FN is a false negative;

FP is a false positive;

$g(I)$ is a method of combining predictions from different networks;

I is an image matrix;

$I_{i,j}$ is a matrix element at position (i, j);

k is a half size of filter window;

M is a width;

N is a height;

n_a is the number of architectures in an ensemble;

n_{one} is a number of networks of the same architecture;

n_e is a number of ensembles;

n_n is a number of networks in an ensemble;

DOI 10.15588/1607-3274-2024-4-10

P is a sequence of ground truth masks;
p_i is a prediction of i-th model;
S is a sequence of predictions of different networks;
TP is true positive;
v is a value of pixel;
v_m is a value of one metric out of three;
X is a pixel sequence;
x is an input image;
σ is a standard deviation that determines the degree of blurring;
σ_s is a degree of spatial smoothing;
σ_r is an intensity smoothing degree;
~*mask* is an inverted mask.

INTRODUCTION

Automation of the spine segmentation process can significantly improve automatic diagnostics of diseases that require precise vertebrae selection. In the absence of a radiologist or his workload, a doctor who needs a radiologist's opinion can make a conclusion himself using the obtained results of automatic segmentation.

A large number of diseases of the spine require a better study of ways to improve its segmentation. The causes of Andersen's lesion are not completely clear, one of the theories accepted today is that primary inflammations are part of ankylosing spondylitis [1]. Cryptococcosis is an infectious disease, the cells of development of which can be vertebrae [2]. Fractures at the level of the 3rd and 4th vertebrae can injure the esophagus. As a result of ankylosing spondylitis, there is also a possibility of damage to the esophagus [3].

By using neural network ensembles, it is possible to improve the accuracy of X-ray image segmentation. Since the quality of X-ray images depends on many factors, this complicates the segmentation process. Analysis and comparison of different neural network ensemble methods can help determine the statistical pattern and select the best algorithm.

The object of study is the process of constructing ensembles of neural networks for segmenting the vertebral region. Different prediction-averaging algorithms are used to construct an ensemble of neural networks. The use of different averaging methods gives different results, which affects the result of using ensembles. The use of ensembles can cause an ambiguous effect, which in turn leads to the need to study these methods for segmenting the lumbar region since different methods have different effects on different objects.

The subject of study is methods of averaging ensembles of neural networks and preprocessing methods. Existing ensemble methods are considered in the context of certain tasks. The use of averaging algorithms provides different results for different objects. The study of existing methods and the development of new ones for segmentation of the vertebral region is a necessary part of the field of studying the segmentation of a certain region of the spine to compare their behavior with methods in another area.

The purpose of the work is to study the effect of preprocessing of vertebral images and compare neural network ensemble methods to determine the best one, the use of which will provide a guaranteed improvement in the segmentation of medical images.

1 PROBLEM STATEMENT

For segmentation of X-ray images, images of different quality are used. The images used can be represented as I, a two-dimensional matrix with the size $M \times N$. Each pixel of the image $I_{i,j}$ has a value from 0 to 255.

An ensemble of neural networks consists of n networks. The creation of ensembles implies the use of neural networks of varying accuracy, which in turn, with a large number of neural networks, worsens the accuracy of segmentation. To solve this problem, it is necessary to supplement existing methods for obtaining ensemble results with new methods or modify existing ones. There are various methods for combining the results of neural network predictions, the most famous of which is the pixel averaging method. For each method that combines ensemble predictions, images are required:

$$S = \{I_i\}_{i=1}^{n}.$$

Due to the problems faced by various ensemble methods, it is necessary to define a method for which the following condition will be satisfied:

$$acc(g_i(S), P) > acc(g_j(S), P),$$

where $i, j \in \{1,2...,n\}$ and $i \neq j$.

2 REVIEW OF THE LITERATURE

Vertebral segmentation plays an integral role in disease diagnosis, preoperative preparation, and subsequent observation. Different quality of X-ray images requires the development of different approaches to extracting the necessary features from images. One of the reasons for obtaining poor-quality images is artefacts in the form of improper exposure. One study has shown that improper breathing technique significantly affects the quality of X-ray images [4]. In different countries, radiologists have different attitudes towards poor-quality images and images with poor image criteria. Non-compliance with image criteria, as well as poor image quality, are reasons for rejecting this image [5]. The dilemma of the impact of radiation on the patient and improving image quality forces doctors to make different decisions. After all, an increase in the dose entails an improvement in image quality, but at the same time, the risk for the patient increases. This issue creates the problem of finding a compromise to select the required radiation dose [6].

The use of machine learning methods for medical image segmentation is a very popular and in-demand approach nowadays. The use of this technique has signifi-

DOI 10.15588/1607-3274-2024-4-10

cantly helped to improve the quality of identifying various types of objects in X-ray images. In the process of improving neural networks and their modernization, various approaches have been proposed, one of which is an ensemble of neural networks. Thanks to ensembles of neural networks, it has become possible to combine less accurate classifiers, resulting in more accurate classifiers [7]. Deep learning models have one main problem – this is the need for a large amount of data and setting up optimal hyperparameters to achieve minimal error. The article [8] discusses the use of several ensemble methods for further application in a wide range of areas. Tests conducted in the article [9] showed that the resulting decrease in Boosting performance was due to overtraining in the presence of noise, which negatively affects the averaging result.

When considering ensembles of neural networks as an approach to improving the quality of segmentation, we should not forget about the methods that are based on the use of only one network. Studies of which have proven their ability to improve the search for specific objects in images. For example, a two-stage method using positioning of lumbar vertebrae and their subsequent segmentation due to the use of several networks at different stages: U-Net and XUnet, showed good performance [10]. The choice between using one network or several is not always obvious. In the article [11], the authors use only one network, since they claim that this approach tries to minimize redundancy in order to reduce the complexity of the network and reduce the training time. They note that this result can be obtained without sacrificing accuracy. The use of a two-branch multi-scale attention module that extracts the necessary information needed for segmentation of the vertebrae and the selection of key information in feature maps was proposed in the article [12]. In the paper [13], the authors proposed a vertebral segmentation method that segments one row of vertebrae as one individual spine object without training data using only manual identification for at least one vertebra. In the second step, they merge the shape prior to the segmentation flow of individual vertebrae. A method for localizing the lumbar spine using YOLOv5 and then passing the localized vertebrae through HED-U-Net to obtain the vertebrae and their edges was proposed in the paper [14].

Most of the studies on medical image segmentation use either neural networks specially designed for a specific task or general-purpose medical networks such as U-Net. The use of U-Net and deep learning for segmentation of the lumbar spine in MRI images demonstrated high segmentation accuracy [15]. Combining spinal canal segmentation using deep learning and morphological operations to solve the redundancy problems and improve the segmentation accuracy was proposed in the paper [16]. As can be seen from the extensive use of single neural networks for vertebrae segmentation, which shows good results, the use of ensembles of neural networks can further improve the results already obtained.

3 MATERIALS AND METHODS

Any machine learning method starts with data preparation. In some studies, on image segmentation you can find a stage of initial image preparation which may include: histogram equalization, median filter, Gaussian filter, etc. To begin with, it is worth considering these methods, because, in the case of medical X-ray images, they may not provide a positive result, but on the contrary, provide a negative result. Since the use of the same Gaussian filter can lead to deterioration in the quality of the boundaries.

The first method worth considering is the well-known histogram equalization algorithm. First, we need to build an initial histogram that will count the number of pixels for each intensity:

$$H(i) = \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} = \begin{cases} 1, I(x,y) = i \\ 0, otherwise \end{cases}.$$

The next step is to normalize the histogram:

$$H_{norm}(i) = \frac{H(i)}{M \times N}.$$

The third stage is the definition of the cumulative function:

$$CDF(i) = \sum_{j=0}^{i} H_{norm}(j).$$

Function to align each pixel:

$$I_{new}(x,y) = 255 \times CDF(I(x,y)).$$

Finally, we need to align all the pixels:

$$I_{new} = \{I_{new}(x,y) | 0 \le x < M, 0 \le y < N\}.$$

The second method for improving the contrast of CLAHE. The adaptive method differs from the usual one in that it calculates several histograms at once, each of the histograms corresponds to a separate section of the input image.

First of all, the core of the block is determined $B_M \times B_N$. Then for each block *b(i,j)* need to create a histogram $H_{i,j}(v)$. To prevent excessive contrast enhancement for each histogram, each block is limited by a threshold *T*. This is done in order to redistribute all values above a given threshold among the remaining values: $H'_{i,j}(v) = \min(H_{i,j}(v), T)$.

The next step is to determine the cumulative function:

$$CDF_{i,j}(v) = \sum_{n=0}^{v} H'_{i,j}(n).$$

DOI 10.15588/1607-3274-2024-4-10

And at the very end, we need to update the pixel values:

$$I_{new}(x,y) = round\left(\frac{CDF_{i,j}(I(x,y)) - \min(CDF_{i,j})}{(B_m \times B_n) - \min(CDF_{i,j})} \times 255\right).$$

The third method is the logarithmic transformation method. This method is used to highlight details with a low contrast level. The method looks like this:

$$I_{new}(x,y) = c \cdot \log(1 + I(x,y)).$$

The fourth method is the median filter. This method is designed to remove unwanted noise by dividing the image into windows in which all pixel values are grouped, after which the median value is determined. First, we need to define a window W with the size $M \times N$, after which it is necessary to extract the values of all pixels from this window:

$$X = \{I(x_i, y_j) | 1 \le i < M, 1 \le j < N\}.$$

After which we will get a sorted array in ascending order X_{sorted}. Let's extract the median value $Med = median(X_{sorted})$, after which we write down the new value $I_{new}(x,y) = Med$.

The fifth method is Gaussian filter, which allows us to remove noise from an image by blurring it. First, the kernel is determined:

$$G(x,y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right).$$

The value of the new pixel after applying this filter can be written as:

$$I_{new}(x,y) = \sum_{i=-k}^{k} \sum_{j=-k}^{k} G(i,j) \cdot I(x+i, y+j).$$

Bilateral filter is used in image processing to remove noise and smooth the image. The filter is calculated using the following formula:

$$I_{new}(x,y) = \frac{1}{W(x,y)} \sum_{i=-k}^{k} \sum_{j=-k}^{k} I(x+i, y+j) \cdot G_S(i,j) \cdot G_r(I(x+i, y+j) - I(x,y)).$$

Gaussian distribution:

$$G_S(i,j) = \exp\left(-\frac{i^2 + j^2}{2\sigma_r^2}\right).$$

Gaussian distribution is also used:

$$G_r(\Delta I) = \exp\left(-\frac{(\Delta I)^2}{2\sigma_r^2}\right).$$

The normalization coefficient has the following form:

$$W(x,y) = \sum_{i=-k}^{k} \sum_{j=-k}^{k} G_S(i,j) \cdot G_r(I(x+i, y+j) - I(x,y)).$$

After it became possible to obtain good results using image segmentation with machine learning, a new method was proposed – an ensemble of neural networks. By combining the predictions of several networks, it became possible to improve the performance of the model. One of the most famous methods that is used to combine the predictions of neural networks is the averaging method. This method has the following form:

$$Mean(n_n) = \frac{1}{n_n} \sum_{i=1}^{n_n} p_i.$$

The next ensemble method that was used was the weighted averaging method. The main difference from the averaging method is that each model is assigned a weight – this is a certain coefficient that is multiplied by the predictions in order to strengthen or weaken the final result of the network. This coefficient can be selected by different features, for example: DICE, IoU, and Recall. The formula for weighted averaging:

$$Weighted = \frac{\sum_{i=1}^{n_n} w_i p_i}{\sum_{i=1}^{n_n} w_i}.$$

If using the averaging method can give a better result, then we can also assume that if we create an averaging method for already averaged results, we can improve the indicators even more. The idea of this method is that we need n results of averaging different ensembles, after which we average these results using the following formula:

$$CumulativeMean = \frac{1}{n_e} \sum_{i=1}^{n_e} Mean_i(n_n).$$

The last prediction fusion method worth considering is the proposed mask shape averaging method. It is based on the idea of transforming the distance of n masks. For each mask, the distance from each pixel of the binary image to the nearest zero pixel is calculated. For zero pixels, the distance will be zero. The following formula displays the distance calculation for one mask:

$$d = dt(mask) - dt(\sim mask).$$

General formula for calculating the distance for all masks:

$$d_{all} = \sum_{i=1}^{n} d(mask_i).$$

DOI 10.15588/1607-3274-2024-4-10

In order to obtain the resulting mask, we sum up the obtained distances and if the sum of the values is greater than 0, then the pixel acquires the values 1 and 0 otherwise:

$$mask_{res} = d_{all} > 0.$$

4 EXPERIMENTS

For this study, open-source images were used [17]. The dataset consisted of 1098 chest X-ray images in lateral projection. The X-ray images depicted men and women. 1020 images were selected for training, the remaining 78 images were used for testing. There were 183 original images, and the remaining 915 were the result of augmentation. The following augmentation operations were performed: 1) random rotation in degrees [–15, 15]; 2) random shift in percentage vertically and horizontally [–10, 10]; 3) random scaling in percent [0.8, 1.2]; 4) random change in brightness [0.8, 1.2]; 5) random change of contrast [0.75, 1.5]. All images were reduced to the same resolution of 512×512 pixels. Figure 1 demonstrates an example of the image used and its masks that were in the dataset.

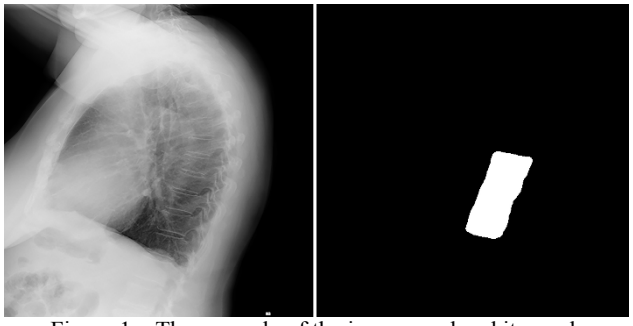

Figure 1 – The example of the image used and its mask

To study the diverse behavior of ensembles, it was decided to train ten models to study their behavior. The list of models used: Fcn8Mobilenet, Fcn8Resnet50, Fcn8Vgg, MobilenetUnet, Pspnet50, Pspnet101, Resnet50Pspnet, Resnet50Segnet, Resnet50Unet, VggUnet. Each used neural network architecture was trained five times. The number of epochs was set to 150, the batch size was eight, and EarlyStopping was used to prevent overfitting, which stopped training if the validation loss value did not improve over ten epochs.

To compare the predictions of neural networks and true masks, the Dice-Sørensen coefficient was used. Which serves as a binary measure of similarity and is expressed by the formula:

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|}.$$

Task 1. To study the meaning of using preprocessing methods, five Fcn8Mobilenet neural networks were used. Different filters were applied to each test image: histogram equalization, CLAHE, logarithmic transformation, median filter, Gaussian filter, and bilateral filter. After one of the above filters was applied to the image, a prediction was made for it from five networks.

Task 2. First, the averaging method was investigated. The main goal was to determine in which cases the best result can be achieved, due to the combination of which networks the model performance can be improved. First, ensembles of one architecture with different numbers of networks in the ensemble were used. Then, the possibility of combining different architectures with different numbers of networks in ensembles was considered. In the first case, for ensembles of one architecture, the following number of networks in one ensemble was used: 2, 3, 4, 5. In the second case of a combination of different architectures, the number of networks of one architecture in the ensemble was also: 2, 3, 4, 5, but the total number of networks in the ensemble was calculated as follows: $A=n_a \times n_{one}$. It is assumed that combining different architectures in an ensemble can have a good effect on the segmentation process due to different feature extraction and their combination. Each architecture will be assigned a code name to make it easier to write down combinations of names in the table. Architecture code names: A – Fcn8Mobilenet, B – Fcn8Resnet50, C – Fcn8Vgg, D – MobilenetUnet, E – Pspnet50, F – Pspnet101, G – Resnet50Pspnet, H – Resnet50Segnet, I – Resnet50Unet, J – VggUnet. Due to the large number of variations of network combinations, all results cannot be displayed in the table, so only a part will be displayed.

Task 3. The weighted averaging method requires a careful selection of weights for each network. In this case, it was proposed to use the following metrics: DICE, Precision, and Recall. Precision indicates the proportion of positive predictions that were specified correctly from all positive cases. The Precision formula is as follows:

$$Precision = \frac{TP}{TP + FP}.$$

The third metric, Recall, shows how many true positive cases were correctly predicted:

$$Recall = \frac{TP}{TP + FN}.$$

The weight for each network was obtained as follows:

$$w = \frac{v_m^{10}}{\sum_{i=1}^{n} v_m^{10}}.$$

Task 4. To study the cumulative averaging method, ten architectures were used, for each architecture four ensembles were used. The result of this algorithm was the combination of the results of the four ensembles. For each cumulative ensemble, averaging ensembles with the same number of networks were used, namely: an ensemble of

DOI 10.15588/1607-3274-2024-4-10

two networks, an ensemble of three networks, an ensemble of four networks and an ensemble of five networks.

Task 5. To implement this method, the training sample was divided into five parts. After that, five Fcn8Mobilenet networks were trained. Each network was trained on different data, so all networks had different images during the training process, which should have expanded the capabilities of the model for image segmentation. After the networks were trained, they were tested in four ensembles.

Task 6. In all previous studies in this article, a maximum of five networks of the same architecture were used, in this method it is proposed to check the influence of a large number of up to 100 networks of the same architecture on the segmentation result. Fcn8Mobilenet was trained 100 times on the same data.

Task 7. The last method that was used to combine the predicted masks was the method of averaging the shapes of masks based on the transformed distance. For its study, ten architectures were used and ensembles of 2, 3, 4, and 5 networks were created. Visually, the work of the algorithm can be seen in Figure 2. Where the white mask is the true mask of the image, the multi-colored contours are the contours of the masks used for averaging, the gold contour is the contour of the averaged mask.

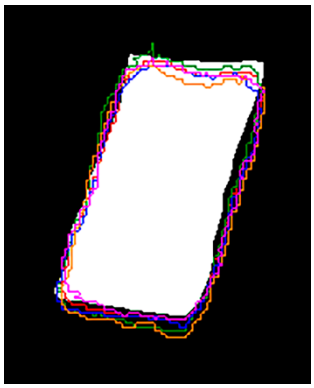

Figure 2 – The example of the method of averaging mask shapes with distance transformation

5 RESULTS

All trained neural networks are shown in Table 1. For greater clarity when comparing with ensemble methods, maximum, minimum and average values are also added.

The obtained results of the study of the application of image preprocessing methods are shown in Table 2. Only for two cases out of 30 a positive result was obtained, in the remaining cases these methods could not positively affect the image, which in turn did not lead to any improvement in the quality of segmentation, but on the contrary, worsened it. And even in these two cases, the result improved insignificantly. The application of these methods can have a good result in other areas, but in the study of X-ray images, they only harm. These methods blur or make the boundaries less clear, as a result of which the extraction of the necessary features suffers.

It was assumed that with the increase in the number of networks in the ensemble, DICE should increase, but this did not happen. Moreover, in most cases, with the increase in the number of networks, the DICE value decreased. Table 3 shows that in 12 cases out of 40 using averaging, it was possible to achieve a positive result compared to single networks. Only in the case of Resnet50Pspnet was it possible to achieve a positive result for all four ensembles. Fcn8Mobilenet was next with three positive results, and Pspnet50 closes the list with two successful ensembles, all other positive architectures had only one ensemble.

Using different architectures to build ensembles gave a positive result in the case of using nine and ten architectures simultaneously, which is demonstrated in Table 4. Combining a smaller number of networks in an ensemble had different results, and to achieve good results, it was necessary to use a large number of ensembles. It can be said with confidence that the simultaneous use of networks with a DICE difference greater than 0.05 will not bring a good result. Otherwise, a positive result could be seen with both a small and large difference in this value.

Table 5 shows the results of weighted averaging, in the process of studying which 90 cases were identified, using this method 22 positive results were obtained compared to the best single networks of the corresponding architectures. The best result was demonstrated by the Resnet50Pspnet architecture, which was able to achieve seven positive results out of nine, which is an excellent result.

The study of the use of the cumulative averaging method revealed that this method can improve the similarity of predictions and true masks compared to the best results of single networks in five cases out of ten. As for the comparison of this method and the averaging method, this method was able to surpass the competitive method in two cases out of ten. The results can be found in Table 6.

Training five networks of the same architecture on different data sets allowed us to minimize errors due to the ability to respond differently to input data. Using different data sets makes it possible to reduce the correlation of model errors. Table 7 shows the DICE for five trained networks. Studying the data provided in Table 8, we can conclude that in three out of four cases, using such an ensemble had a better result compared to the maximum result of a single network.

The use of a large number of networks of the same architecture in the ensemble is shown in Table 9. The expected increase in the result at each step of increasing the number of networks in the ensemble did not occur, but nevertheless, the best result was achieved using 100 networks. Although it is worth considering that the weight of 100 trained networks was 252 GB.

The last method considered was able to show the best result of all. 40 ensemble cases were considered, in 20 cases this method was better than maximal single networks and in 29 cases this method outperformed the averaging method, this can be seen in Table 10.

DOI 10.15588/1607-3274-2024-4-10

Table 1 – Similarity measure for trained neural networks

Architecture name	Network number					Statistics		
	1	2	3	4	5	Min	Max	Avg
Fcn8Mobilenet	0.9179	0.9142	0.9061	0.9035	0.8933	0.8933	0.9179	0.9070
Fcn8Resnet50	0.9539	0.9397	0.9272	0.9218	0.9188	0.9188	0.9539	0.9323
Fcn8Vgg	0.9457	0.9135	0.8896	0.8885	0.88442	0.8844	0.9457	0.9043
MobilenetUnet	0.9119	0.9078	0.9045	0.8780	0.8755	0.8755	0.9119	0.8956
Pspnet50	0.9338	0.9121	0.9091	0.9040	0.8897	0.8897	0.9338	0.9097
Pspnet101	0.9480	0.9358	0.8978	0.8767	0.8446	0.8446	0.9480	0.9006
Resnet50Pspnet	0.8938	0.8933	0.8920	0.8865	0.8553	0.8553	0.8938	0.8842
Resnet50Segnet	0.9420	0.9154	0.8941	0.8934	0.8788	0.8788	0.9420	0.9048
Resnet50Unet	0.9315	0.9157	0.9120	0.9115	0.9043	0.9043	0.9315	0.9150
VggUnet	0.9151	0.9067	0.8955	0.8849	0.8757	0.8757	0.9151	0.8956

Table 2 – Similarity measure for image preprocessing results

Method name	Network number					Statistics	
	1	2	3	4	5	Min	Max
HE	0.8998	0.8971	0.8831	0.8871	0.8804	0.8804	0.8998
CLAHE	0.9206	0.8872	0.9037	0.8949	0.8772	0.8772	0.9206
LT	0.9108	0.9041	0.8771	0.9094	0.8595	0.8595	0.9108
MF	0.8997	0.8928	0.8890	0.8894	0.8860	0.8860	0.8997
GF	0.9156	0.9023	0.9026	0.9003	0.8933	0.8933	0.9156
BF	0.8730	0.8795	0.8684	0.8676	0.86062	0.8606	0.8795

Table 3 – Ensembles of neural networks of the same architecture

Architecture name	Number of networks in the ensemble				Statistics	
	2	3	4	5	Min	Max
Fcn8Mobilenet	0.9194	0.9192	0.9136	0.9185	0.9136	0.9194
Fcn8Resnet50	0.9449	0.9493	0.9449	0.9466	0.9449	0.9493
Fcn8Vgg	0.9312	0.9372	0.9260	0.9280	0.9268	0.9372
MobilenetUnet	0.9104	0.9144	0.9057	0.9047	0.9047	0.9144
Pspnet50	0.9208	0.9407	0.9360	0.9324	0.9208	0.9407
Pspnet101	0.9382	0.9513	0.9330	0.9474	0.9330	0.9513
Resnet50Pspnet	0.8943	0.9090	0.9062	0.9049	0.8943	0.9090
Resnet50Segnet	0.9261	0.9352	0.9295	0.9289	0.9261	0.9352
Resnet50Unet	0.9261	0.9352	0.9295	0.9289	0.9261	0.9352
VggUnet	0.9062	0.9124	0.9146	0.9147	0.9062	0.9147

Table 4 – Ensembles of neural networks of different architectures

Combinations	Number of networks of same architecture in ensemble					Statistics	
	1	2	3	4	5	Min	Max
A+B	0.9337	0.9383	0.9440	0.9399	0.9411	0.9383	0.9440
A+B+C	0.9524	0.9494	0.9467	0.9443	0.9432	0.9432	0.9524
A+B+C+D	0.9480	0.9468	0.9449	0.9423	0.9399	0.9399	0.9480
A+B+C+D+E	0.9496	0.9485	0.9470	0.9447	0.9428	0.9428	0.9496
A+B+C+D+E+F	0.9510	0.9510	0.9490	0.9473	0.9458	0.9458	0.9510
A+B+C+D+E+F+G	0.9525	0.9496	0.9484	0.9451	0.9437	0.9437	0.9525
A+B+C+D+E+F+G+H	0.9519	0.9509	0.9478	0.9466	0.9449	0.9449	0.9519
A+B+C+D+E+F+G+H+I	0.9549	0.9513	0.9490	0.9472	0.9463	0.9463	0.9549
A+B+C+D+E+F+G+H+I+J	0.9541	0.9512	0.9489	0.9474	0.9464	0.9464	0.9541
A+D	0.9138	0.9216	0.9269	0.9247	0.9210	0.9138	0.9269
A+E	0.9278	0.9321	0.9331	0.9348	0.9387	0.9278	0.9387
A+G	0.9042	0.9164	0.9222	0.9237	0.9240	0.9042	0.9240
A+J	0.9315	0.9323	0.9361	0.9311	0.9312	0.9311	0.9361
A+K	0.9177	0.9253	0.9248	0.9216	0.9240	0.9177	0.9253
C+E	0.9471	0.9450	0.9465	0.9441	0.9335	0.9335	0.9471
C+F	0.9529	0.9493	0.9447	0.9429	0.9431	0.9429	0.9529
C+H	0.9468	0.9467	0.9427	0.9433	0.9416	0.9416	0.9468
C+J	0.9519	0.9429	0.9458	0.9421	0.9378	0.9378	0.9519
D+E+J	0.9445	0.9402	0.9401	0.9413	0.9395	0.9395	0.9445
D+E+K	0.9368	0.9397	0.9388	0.9360	0.9354	0.9354	0.9397
D+E+G	0.9388	0.9369	0.9373	0.9329	0.9316	0.9316	0.9388
D+E+G+J	0.9407	0.9428	0.9434	0.9413	0.9393	0.9393	0.9434
D+E+G+K	0.9373	0.9388	0.9373	0.9358	0.9345	0.9345	0.9388
D+K+E+J	0.9393	0.9425	0.9423	0.9418	0.9417	0.9393	0.9425

DOI 10.15588/1607-3274-2024-4-10

Table 5 – Weighted averaging

Architecture name	Results								
	Precision			Recall			DSC		
	N_n								
	3	4	5	3	4	5	3	4	5
Fcn8Mobilenet	0.9141	0.9181	0.9185	0.9141	0.9168	0.9169	0.9141	0.9181	0.9185
Fcn8Resnet50	0.9308	0.9465	0.9466	0.9308	0.9465	0.9466	0.9308	0.9465	0.9466
Fcn8Vgg	0.9312	0.9403	0.9385	0.9132	0.9333	0.9280	0.9312	0.9403	0.9385
MobilenetUnet	0.9144	0.9144	0.9144	0.9031	0.9118	0.9047	0.9031	0.9118	0.9047
Pspnet50	0.8981	0.9372	0.9324	0.8981	0.9378	0.9329	0.8980	0.9372	0.9324
Pspnet101	0.9382	0.9382	0.8726	0.9520	0.9513	0.8723	0.9382	0.9504	0.8722
Resnet50Pspnet	0.8860	0.9074	0.9049	0.8943	0.9082	0.9050	0.8860	0.9074	0.9049
Resnet50Segnet	0.8983	0.8939	0.9095	0.9321	0.9321	0.9321	0.9338	0.9399	0.9405
Resnet50Unet	0.9142	0.9285	0.9289	0.9142	0.9285	0.9289	0.9143	0.9285	0.9289
VggUnet	0.9162	0.9127	0.9130	0.8952	0.9118	0.9181	0.8952	0.9084	0.9147

Table 6 – Cumulative averaging

Score	Architecture code name									
	A	B	C	D	E	F	G	H	I	J
DSC	0.9193	0.9490	0.9379	0.9088	0.9348	0.9504	0.9054	0.9324	0.9296	0.9164

Table 7 – Fcn8Mobilenet trained on 5 different datasets

Score	Network number					Statistics		
	1	2	3	4	5	Min	Max	Avg
DSC	0.9004	0.8441	0.8607	0.8881	0.8773	0.8441	0.9004	0.8741

Table 8 – The result of averaging Fcn8Mobilenet trained on 5 datasets

Score	Number of networks in the ensemble				Statistics		
	2	3	4	5	Min	Max	Avg
DSC	0.9023	0.8789	0.9023	0.9006	0.8789	0.9023	0.8960

Table 9 – Result of averaging a large number of Fcn8Mobilenet in one ensemble

Score	Number of networks in the ensemble										Statistics		
	10	20	30	40	50	60	70	80	90	100	Min	Max	Avg
DSC	0.9240	0.9252	0.9242	0.9251	0.9254	0.9259	0.9257	0.9250	0.9255	0.9261	0.9185	0.9261	0.9246

Table 10 – The result of applying the mask shape averaging method

Architecture name	Number of networks in the ensemble				Statistics		
	2	3	4	5	Min	Max	Avg
Fcn8Mobilenet	0.925156	0.92502	0.920583	0.922195	0.920583	0.925156	0.923239
Fcn8Resnet50	0.951575	0.945714	0.944762	0.944422	0.944422	0.951575	0.946618
Fcn8Vgg	0.938029	0.930131	0.93025	0.927827	0.927827	0.938029	0.931559
MobilenetUnet	0.913377	0.91427	0.910609	0.906647	0.906647	0.91427	0.911226
Pspnet50	0.934716	0.933914	0.935697	0.931073	0.931073	0.935697	0.93385
Pspnet101	0.94993	0.944895	0.93448	0.856578	0.856578	0.94993	0.921471
Resnet50Pspnet	0.905457	0.909698	0.912497	0.906793	0.905457	0.912497	0.908611
Resnet50Segnet	0.935092	0.93356	0.932842	0.92569	0.92569	0.935092	0.931796
Resnet50Unet	0.937276	0.937114	0.931678	0.929995	0.929995	0.937276	0.934016
VggUnet	0.919379	0.915692	0.912555	0.916767	0.912555	0.919379	0.916098

6 DISCUSSION

Having considered the comparison of different pre-processing methods, we can conclude that the use of such methods is undesirable for the tasks of segmentation of chest X-ray images. The use of these methods did not produce the desired result, but on the contrary, made it worse. The use of these methods reduced the visibility of boundaries, blurred the image, and some methods even increased noise. The presence of these methods in the segmentation algorithm complicates the extraction of the necessary features from images and is definitely not recommended.

The use of averaging method is the simplest ensemble method, which makes it easy to use. However as the result showed, only in 12 cases out of 40 could an improvement be achieved compared to using single networks. This method, in the presence of a large number of predictions of poor quality, has the property of deteriorating the resulting prediction. The use of this method in problems that require guaranteed accuracy is ambiguous and undesirable. At the same time, the study of this method proved that the use of ensembles of different architectures has an advantage over the use of ensembles of the same architectures due to the combination of different data extraction methods.

DOI 10.15588/1607-3274-2024-4-10

The weighted averaging method was able to show improvement over single networks in 22 cases out of 90, suggesting that applying weights to networks would strengthen strong networks and weaken weak networks. However, this method did not show superiority over the averaging method. As can be seen in Table 5, the weights for this method were determined based on three metrics. The result of this algorithm directly depends on the selection of the necessary weights, so further in-depth study of this method can improve its accuracy.

The cumulative method, based on the idea of averaging already averaged predictions, was expected to be better than averaging method. But this is only guaranteed if averaging is able to provide only positive results. Only in five out of ten cases compared to single networks and in two out of ten compared to ensembles of averaging did this method show an increase in results. Such results indicate that the use of averaging is unstable and further highlight the controversy over the use of averaging.

Using an ensemble of networks of the same architecture trained on different datasets helps to reduce the correlation of model errors. In three out of four cases, using ensembles with the usual averaging method was better than using a single network with better accuracy. Although the obtained result was small, it should not be forgotten that the used dataset also consists of augmented data and if the opportunity to use a dataset with a large number of unique images and its division into subsets was provided, the result could have been significantly better.

Using 100 trained networks showed the same ambiguous result. Although the ensemble of 100 networks was the most successful in comparison with other numbers of networks in the ensemble, the total weight of 252 GB of all networks is hard to imagine in use in a medical institution.

The method of averaging mask shapes using distance transformation was able to demonstrate a good result. This method was better in 20 cases compared to single networks, while conventional averaging brought a result of 12. It managed to show a result better than averaging, because the algorithm is based not on averaging pixel values, but on calculating mask distances. But this method has one nuance that stops it on the way to a complete improvement of the results. This method is poorly susceptible to artefacts in the form of growths on the mask, since the shapes of masks are averaged, any mask used whose shape will differ significantly from the shape of the desired object crosses out all the positive aspects of its work. To eliminate this interference, it is worth using methods for removing artifacts on masks which will definitely lead to an improvement in the work of this algorithm.

CONCLUSIONS

In this study, different methods of combining neural network ensemble predictions for segmentation of the thoracic spine region were examined. The extensive study of different methods allowed us to further explore the choice and advantage of specific methods. In any case, ensemble methods were able to demonstrate improvement in segmentation, although not in all cases considered.

The impact of image preprocessing on X-ray image segmentation tasks was also studied. The obtained results gave reason to doubt the appropriateness of these methods for solving such problems.

The scientific novelty of the obtained results is the effect of such neural network prediction fusion algorithms using such neural network architectures was examined for the first time, and the effect of six image preprocessing methods for segmentation was studied. This allows us to select the appropriate method for further spine segmentation studies.

The practical significance consists in the fact that a comparison of several methods for combining mask predictions on chest X-ray images in the lateral projection was made, which made it possible to apply this approach to creating automatic segmentation of vertebrae or necessary areas of the spine and implement them in medical institutions.

Prospects for further research is a more detailed study of the weighted averaging method with a more extensive selection of weights that may depend on a large number of metrics or other parameters. Also, further improvement of the mask shape averaging method may bring a more successful result than was obtained.

ACKNOWLEDGEMENTS

We express our gratitude to A. Pidhornyi Institute of Power Machines and Systems of NAS of Ukraine for the opportunity to conduct this study.

REFERENCES

1. Wu H., Wu X., Wu T., Miao X., Zheng S., Huang G., Cheng X. Detection Ewingella americana from a patient with Andersson lesion in ankylosing spondylitis by metagenomic next-generation sequencing test: a case report, *BMC Musculoskelet Disord,* 2024, Vol. 25, P. 568. DOI: 10.1186/s12891-024-07680-y
2. Zhou Y., Huang X., Liu Y., Zhou X., Liu Q. Destructive Cryptococcal Osteomyelitis Mimicking Tuberculous Spondylitis, *American Journal of Case Reports,* 2024, Vol. 25, P. e944291. DOI: 10.12659/AJCR.944291
3. Smolle M. A., Maier A., Lindenmann J., Porubsky C., Leithner J., Smolle-Juettner F. M. Esophageal perforation with near fatal mediastinitis secondary to Th3 fracture, *Wiener klinische Wochenschrift,* 2024. DOI: 10.1007/s00508-024-02397-3
4. Sønderby A. H., Thomsen H., Skals R. G., Storm S., Leutscher P.D.C., Simony A. Thoracic spine X-ray examination of patients with back pain using different breathing technique and exposure times – A diagnostic study, *Radiography,* 2024, Vol. 30, pp 582–288. DOI: 10.1016/j.radi.2024.01.011
5. Kjelle E., Chilanga C. The assessment of image quality and diagnostic value in X-ray images: a survey on radiographers' reasons for rejecting images, *Insights into imaging*, 2022, Vol. 13(1), № 36. DOI: 10.1186/s13244-022-01169-9
6. Ullman G. Quantifying image quality in diagnostic radiology using simulation of the imaging system and model observers. Linköping, Sweden, 2008, 85 p.
7. Dietterich T. G. Ensemble Methods in Machine Learning, *Multiple Classifier Systems. MCS 2000. Springer. Berlin,*

DOI 10.15588/1607-3274-2024-4-10

Heidelberg, 2000, pp. 1–15. (Lecture Notes in Computer Science, Vol. 1857). DOI: 10.1007/3-540-45014-9_1
8. Mohammed A., Kora R. A comprehensive review on ensemble deep learning: Opportunities and challenges, *Journal of King Saud University – Computer and Information Sciences*, 2023, Vol. 35(2), pp. 757–774. DOI: 10.1016/j.jksuci.2023.01.014
9. Maclin R., Opitz D. W. Popular Ensemble Methods: An Empirical Study, *Journal of Artificial Intelligence Research*, 1999, Vol. 11, pp. 169–198.
10. Lu H., Li M., Zhang Y., Yu L. Lumbar spine segmentation method based on deep learning, *Journal of applied clinical medical physics*, 2023, Vol. 24, Iss. 6, P. e13996. DOI: 10.1002/acm2.13996
11. Xiong X. Graves S. A., Gross B. A., Buatti J. M., Beichel R. R. Lumbar and Thoracic Vertebrae Segmentation in CT Scans Using a 3D Multi-Object Localization and Segmentation CNN, *Tomography*, 2024, Vol. 10(5), pp. 738–760. DOI: 10.3390/tomography10050057
12. Li H., Luo H., Huan W., Shi Z., Yan C., Wang L., Mu Y., Liu Y. Automatic lumbar spinal MRI image segmentation with a multi-scale attention network, *Neural computing & applications*, 2021, Vol. 33(18), pp. 11589–11602. DOI: 10.1007/s00521-021-05856-4
13. Khandelwal P., Collins L. D., Siddiqi K. Spine and Individual Vertebrae Segmentation in Computed Tomography Images Using Geometric Flows and Shape Priors, *Frontiers in Computer Science*, 2021, Vol. 3. DOI: 10.3389/fcomp.2021.592296
14. Mushtaq M., Akram M. U., Alghamdi N. S., Fatima J., Masood R. F. Localization and Edge-Based Segmentation of Lumbar Spine Vertebrae to Identify the Deformities Using Deep Learning Models, *Sensors*, 2022, Vol. 22(4), P. 1547. DOI: 10.3390/s22041547
15. Liang Y., Fang Y. T., Lin T. C., Yang C. R., Chang C. C., Chang H. K., Ko C. C., Tu T. H., Fay L. Y., Wu J. C., Huang W. C., Hu H. W., Chen Y. Y., Kuo C. H. The Quantitative Evaluation of Automatic Segmentation in Lumbar Magnetic Resonance Images, *Neurospine*, 2024, Vol. 21(2). pp. 665–675. DOI: 10.14245/ns.2448060.030
16. Zhou Z., Wang S., Zhang S., Pan X., Yang H., Zhuang Y., Lu Z. Deep learning-based spinal canal segmentation of computed tomography image for disease diagnosis: A proposed system for spinal stenosis diagnosis, *Medicine*, 2024, Vol. 103(18), P. e37943. DOI: 10.1097/MD.0000000000037943
17. Vindr.ai Datasets: SpineXR. [Electronic resource]. Access mode: https://vindr.ai/datasets/spinexr

Received 02.09.2024.
Accepted 24.10.2024.

УДК 004.8

ВПЛИВ ПОПЕРЕДНЬОЇ ОБРОБКИ ТА ПОРІВНЯННЯ НЕЙРОМЕРЕЖЕВИХ АНСАМБЛЕВИХ МЕТОДІВ ДЛЯ СЕГМЕНТАЦІЯ ГРУДНОГО ВІДДІЛУ ХРЕБТА НА РЕНТГЕНІВСЬКИХ ЗНІМКАХ

Конюхов В. Д. – аспірант, Інститут енергетичних машин і систем ім. А. М. Підгорного НАН України, Харків, Україна.
Моргун О. М. – канд. фіз.-мат. наук, директор ТОВ «Лабораторія рентгенівської медичної техніки», Харків, Україна.
Нємченко К. Е. – д-р фіз.-мат. наук, завідувач кафедри Харківський національний університет імені В. Н. Каразіна, Харків, Україна.

АНОТАЦІЯ

Актуальність. Автоматична сегментація медичних знімків відіграє важливу роль у процесі автоматизації визначення захворювань різного роду області хребта, а використання рентгенографії є найдоступнішим засобом передбачення захворювань. За багато років було проведено безліч досліджень на тему сегментації зображень. Одним із багатьох методів покращення сегментації зображень є застосування ансамблів нейронних мереж.

Метою даного дослідження було розглянути вплив попередньої обробки зображень та вивчити і порівняти головні методи ансамблів нейронних мереж та їх вплив на сегментацію області хребта, в даному дослідженні розглядалася область яка складається з хребців: Th8, Th9, Th10, Th11.

Метод. Для початку було розглянуто вплив попередньої обробки рентгенівських зображень, яка включала в себе наступні методи: вирівнювання гістограми для поліпшення контрасту, адаптивне вирівнювання гістограми з обмеженням контрасту, метод логарифмічного перетворення, медіанний фільтр, Гауссово згладжування. Для вивчення впливу ансамблю нейронних мереж на якість сегментації використовувалися такі методи: метод усереднення – найпростіший метод половинного усереднення; зважене усереднення – покращена версія методу усереднення, яка використовує ваги для кожної мережі, чим більша вага мережі – тим більший її вплив на усереднення; метод усереднення усереднених зображень – модифікований метод усереднення в якому кожен ансамбль отримує усереднене зображення, після чого всі результати ансамблів усереднюються; метод усереднення мереж навчених на різних даних – використовується *n* мереж, навчальна вибірка розбивається на *n* частин, кожна нейронна мережа навчається на своїй підмножині даних, в результаті для передбачень використовується звичайний метод усереднення; метод усереднення для великої кількості мереж – у цьому методі було навчено 100 нейронних мереж, після чого використовувався звичайний метод усереднення; метод усереднення контурів – даний метод усереднює всі контури в результаті чого виходить один середній контур.

Результати. Було досліджено, що застосування різних методів попередньої обробки зображень не гарантує поліпшення якості сегментації області хребта на рентгенівських знімках, а навіть навпаки погіршує якість сегментації. Були розглянуті різні методи об'єднання передбачень ансамблів нейронних мереж, що дало можливість дізнатися плюси та мінуси конкретних методів для завдання сегментації рентгенівських знімків.

Висновки. Проведені експерименти дали можливість зробити висновок, що застосування будь-яких методів попередньої обробки не варто використовувати для сегментації рентгенівських знімків. Також завдяки великій кількості архітектур і методів об'єднання передбачень було вивчено поведінку ансамблевих методів що дозволить надалі визначити необхідний підхід для сегментації рентгенівських знімків. Подальше вивчення методу зваженого усереднення і методу усереднення форм масок дасть можливість поліпшити отриманий результат і досягти ще більшого успіху в сегментації.

КЛЮЧОВІ СЛОВА: машинне навчання, розпізнавання образів, нейронна мережа, сегментація зображення, комп'ютерний зір.

DOI 10.15588/1607-3274-2024-4-10

ЛІТЕРАТУРА

1. Detection Ewingella americana from a patient with Andersson lesion in ankylosing spondylitis by metagenomic next-generation sequencing test: a case report / [H. Wu, X. Wu, T. Wu et al.] // BMC Musculoskelet Disord. – 2024. – Vol. 25. – P. 568. DOI: 10.1186/s12891-024-07680-y
2. Destructive Cryptococcal Osteomyelitis Mimicking Tuberculous Spondylitis / [Y. Zhou, X. Huang, Y. Liu et al.] // American Journal of Case Reports. – 2024. – Vol. 25. – P. e944291. DOI: 10.12659/AJCR.944291
3. Esophageal perforation with near fatal mediastinitis secondary to Th3 fracture / [M. A. Smolle, A. Maier, J. Lindenmann et al.] // Wiener klinische Wochenschrift. – 2024. DOI: 10.1007/s00508-024-02397-3
4. Thoracic spine X-ray examination of patients with back pain using different breathing technique and exposure times – A diagnostic study / [A. H. Sønderby, H. Thomsen, R. G. Skals et al.] // Radiography. – 2024. – Vol. 30. – P. 582–288. DOI: 10.1016/j.radi.2024.01.011
5. Kjelle E. The assessment of image quality and diagnostic value in X-ray images: a survey on radiographers' reasons for rejecting images / E. Kjelle, C. Chilanga // Insights into imaging. – 2022. – Vol. 13(1), № 36. DOI: 10.1186/s13244-022-01169-9
6. Ullman G. Quantifying image quality in diagnostic radiology using simulation of the imaging system and model observers / G. Ullman. – Linköping, Sweden, 2008. – 85 p.
7. Dietterich T. G. Ensemble Methods in Machine Learning / T. G. Dietterich // Multiple Classifier Systems. MCS 2000. Springer. – Berlin, Heidelberg, 2000. – P. 1–15. – (Lecture Notes in Computer Science, Vol. 1857). DOI: 10.1007/3-540-45014-9_1
8. Mohammed A. A comprehensive review on ensemble deep learning: Opportunities and challenges / A. Mohammed, R. Kora // Journal of King Saud University – Computer and Information Sciences. – 2023. – Vol. 35(2). – P. 757–774. DOI: 10.1016/j.jksuci.2023.01.014
9. Maclin R. Popular Ensemble Methods: An Empirical Study / R. Maclin, D. W. Opitz // Journal of Artificial Intelligence Research. – 1999. – Vol. 11. – P. 169–198.
10. Lumbar spine segmentation method based on deep learning / [H. Lu, M. Li, Y. Zhang, L. Yu] // Journal of applied clinical medical physics. – 2023. – Vol. 24(6). – P. e13996. DOI: 10.1002/acm2.13996
11. Xiong X. Lumbar and Thoracic Vertebrae Segmentation in CT Scans Using a 3D Multi-Object Localization and Segmentation CNN / [X. Xiong, S. A. Graves, B. A. Gross et al.] // Tomography. – 2024. – Vol. 10(5). – P. 738–760. DOI: 10.3390/tomography10050057
12. Automatic lumbar spinal MRI image segmentation with a multi-scale attention network / [H. Li, H. Luo, W. Huan et al.] // Neural computing & applications. – 2021. – Vol. 33(18). – P. 11589–11602. DOI: 10.1007/s00521-021-05856-4
13. Khandelwal P. Spine and Individual Vertebrae Segmentation in Computed Tomography Images Using Geometric Flows and Shape Priors / P. Khandelwal, L. D. Collins, K. Siddiqi // Frontiers in Computer Science. – 2021. – Vol. 3. DOI: 10.3389/fcomp.2021.592296
14. Localization and Edge-Based Segmentation of Lumbar Spine Vertebrae to Identify the Deformities Using Deep Learning Models / [M. Mushtaq, M. U. Akram, N. S. Alghamdi et al.] // Sensors. – 2022. – Vol. 22(4). – P. 1547. DOI: 10.3390/s22041547
15. The Quantitative Evaluation of Automatic Segmentation in Lumbar Magnetic Resonance Images / [Y. W. Liang, Y. T. Fang, T. C. Lin et al.] // Neurospine. – 2024. – Vol. 21(2). – P. 665–675. DOI: 10.14245/ns.2448060.030
16. Zhou, Z. Deep learning-based spinal canal segmentation of computed tomography image for disease diagnosis: A proposed system for spinal stenosis diagnosis / [Z. Zhou, S. Wang, S. Zhang et al.] // Medicine. – 2024. – Vol. 103(18). – P. e37943. DOI: 10.1097/MD.0000000000037943
17. Vindr.ai Datasets: SpineXR. [Electronic resource]. – Access mode: https://vindr.ai/datasets/spinexr

DOI 10.15588/1607-3274-2024-4-10