The using of machine learning and neural networks in the processing of computer simulation results for medical diagnostics

Maxim Polyakov
Department of information systems and computer modeling
Volgograd State University
Volgograd, Russia
m.v.polyakov@volsu.ru

Alexander Khoperskov
Department of information systems and computer modeling
Volgograd State University
Volgograd, Russia
khoperskov@volsu.ru

Egor Borisovskii
Department of information systems and computer modeling
Volgograd State University
Volgograd, Russia
infomod@volsu.ru

Egor Emelyanov
Department of information systems and computer modeling
Volgograd State University
Volgograd, Russia
infomod@volsu.ru

***Abstract*—We use machine learning technologies and neural networks to improve the efficiency of medical diagnostics based on the microwave radiometry. Originality of our approach is that we use the results of computer modeling temperature fields in multicomponent biological tissues to build training and test data sets. We investigate limits of applicability of the method for diagnosing breast cancer using microwave radiometry data. Task of determining diagnostic quality for different sizes of tumors seems promising to us.**

***Keywords*—computer modeling, machine learning, neural networks, diagnosis of cancer, microwave radiometry**

I. Introduction

Early diagnosis of breast cancer is an important issue in modern medicine. Incidence of breast cancer is increasing worldwide and is one of the most common types of cancer. At the moment, there is no effective means to prevent breast cancer. Probability of successful treatment and full recovery of the patient depends entirely on early detection and diagnosis. Breast cancer is a curable disease with a probability of 97% with early detection [4]. Microwave radiometry is a non-invasive method for measuring internal temperature in biological tissues. This method is based on measuring the self-radiation of the biological tissues in the radio to microwave range. Possibility of using microwave radiometry to diagnose breast cancer was shown in the article [3] for the first time. We can assume that the theoretical basis of microwave radiometry in mammology is based on the research of the French scientist M. Gautherie [6] to some extent. He convincingly showed that the heat release of the tumor is directly proportional to the rate of its growth. His research was based on clinical data from more than 85,000 patients. Microwave radiometry has a unique ability to detect fast-growing tumors in the first place. Microwave radiometry is also used in other fields of medicine [1], [7]. The paper [5] states that the minimum size of a cancerous tumor detected by mammography is 1.68 cm in diameter. The task of microwave radiometry is to detect smaller tumors. Microwave radiometry can also detect cancerous tumors or early structural changes that are not detected and can be skipped when using mammography.

II. Problem statement

Mathematical models and numerical methods that we use to construct a sample of temperature data are described in [9]. We have verified models and it has shown the effectiveness of building models of the mammary glands healthy patients (without cancer pathologies) [10]. The main task of this study is to determine threshold value size of the tumor, which can be detected by microwave radiometry.

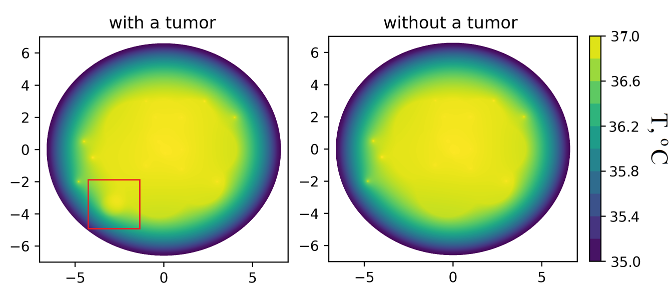

Fig. 1. Distribution of thermodynamic temperature at a depth 4 cm obtained by computer simulation. On the left is a model with a tumor of radius R = 0.5 cm.

To do this, we need to build samples data from computer modeling of breast temperatures, volume which will allow for machine learning, as well as binary classification of test data (healthy-sick). When building a large volume of models, it is necessary to take into account the fact that most often malignant tumors appear in the upper outer quadrant. We present results of numerical simulation thermal dynamics in the biological tissues of the mammary gland (Fig. 1.).

III. Machine learning methods and tools

The following methods are used for binary classification of computer modeling data: support vector machines (SVM), k-nearest neighbors, (KNN) and the naive Bayesian classifier (NBC). These methods are implemented in the Scikit-learn Python library. In addition to the classification task, this library allows you to build regressions, perform clusterization, and so on.

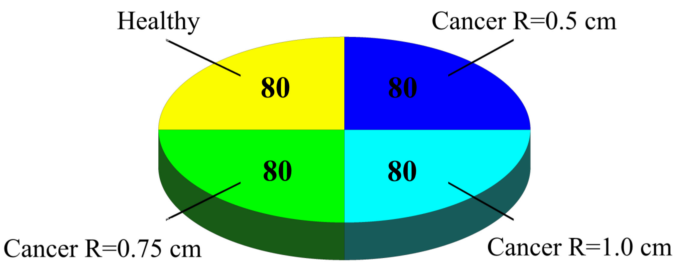

Fig. 2. Structure of source data set.

The training set is a temperature data set at points of the breast (according to the survey method [2]) in the microwave range (depth temperature) and in the infrared range (skin temperature). Each model has the “1” – sick or “0” – healthy flag. Equal data samples containing 80 models were constructed: without a tumor, with a tumor of radii R=0.5 cm, R=0.75 cm and R=1 cm (Fig. 2.). A cross section was made with skin and depth temperatures of points 0,..., 8 and we have a combined data set

$$X = \begin{bmatrix} T_0^1 & T_1^1 & \dots & T_{18}^1 \\ T_0^2 & T_1^2 & \dots & T_{18}^2 \\ \dots & \dots & \dots & \dots \\ T_0^{240} & T_1^{240} & \dots & T_{18}^{240} \end{bmatrix}, Y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_{240} \end{bmatrix}, \quad (1)$$

where $T_0^i, \dots, T_8^i$ are internal temperature at the points $0, \dots, 8$, $T_9^i, \dots, T_{18}^i$ are temperature of the skin at points $0, \dots, 8$ and $y_i \in \{\text{Health}, R\text{=0.5}, R\text{=0.75}, R\text{=1.0}\}$ is label i model.

Ratio of cancer and healthy models in the training and test sets is assumed to be equal, in order to preserve data uniformity. We conducted a statistical analysis of the data at the first stage. It showed general differences between the data sets.

Qualitative analysis of large amounts of data is particularly difficult. Statistical analysis confirms reliable differences in temperature data at certain points of the breast (Fig. 3.).

IV. The using of neural networks

Neural networks are another way to perform binary classification. It is necessary that the input data lie in a single range for successful training of the neural network. We used

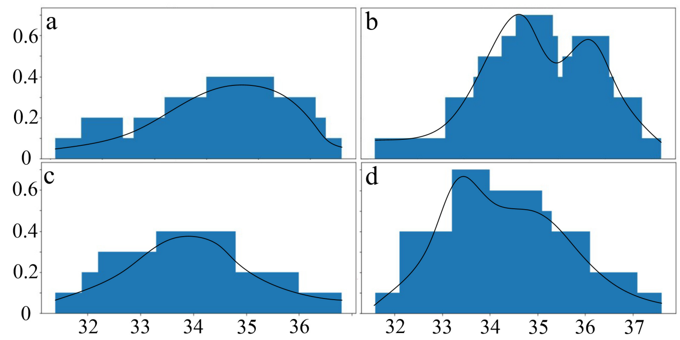

Fig. 3. Frequency distribution of internal temperature data for a) points “0” of the model without a tumor, b) points “0” of the model with a tumor R =0.75 cm, c) points “3” models without a tumor, d) points “3” models with a tumor R =0.75 cm.

the classic standardization method to bring an array of input data to a single view. After applying it, each attribute has an average of “0” and a variance of “1”. Data normalization was performed using the formula

$$\hat{x} = \frac{x_{ij} - M_j}{S_j}, \quad (2)$$

where M_j is the average value of the samples, and S_j is the standard deviation of the samples.

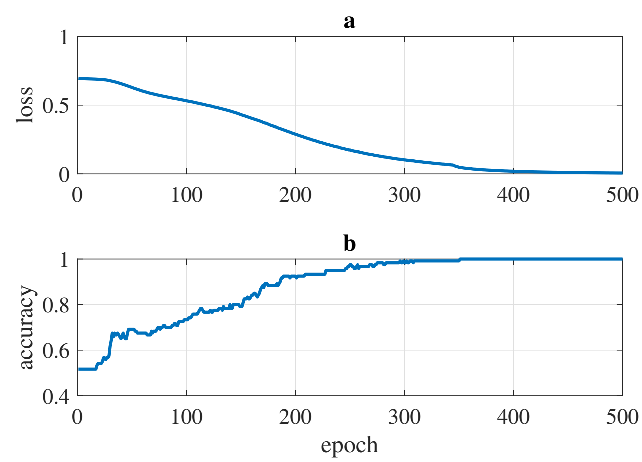

Fig. 4. Dependence loss (a) and accuracy (b) on the number of training epochs of the neural network.

We solve binary classification problem, so number of neurons in the output layer is two. The softmax function is selected as the activation function for last layer. This function converts the network output vector to another vector. The coordinate of this vector is located in range $[0; 1]$. The sum of the coordinates is 1. We define patient class by the largest coordinate in output vector. Parameters of the layers and their number were selected empirically. We were based on the results of testing.

To train a neural network, the original sample consisting of 160 numerical simulation results was randomly mixed and divided into two parts, the first p art c ontaining 1 20 survey results, the second part containing 40. Gradient optimization methods were used to train neural networks.

In order for the model not to be retrained, the number of epochs was set to 500. The Fig.4. shows the accuracy of model and model loss in learning process.

V. Discussion of results

Experiments have shown a significant d ependence o f the effectiveness diagnostics based on microwave radiometry data on size of the tumor.

TABLE I. Effectiveness of data classification methods

	NBC	KNN	SVM
R=0.5 cm	0.475	0.525	0.575
R=0.75 cm	0.7	0.675	0.725
R=1 cm	0.74	0.75	0.79

The support vector method gave the best result in relation to other machine learning methods (Table 1), which indicates that this method is better applicable to this type of tasks and to this structure of training data set. The SVM method's gain in relation to the NBC is 10% for a data set with a tumor R=0.5 cm. Which is significant for the task of medical diagnostics. Ratio of correctly classified models to volume of test sample was used as a measure of efficiency for comparing methods. $E = \frac{\omega}{F}$, where ω is number of correctly recognized models, and F is volume of test data set.

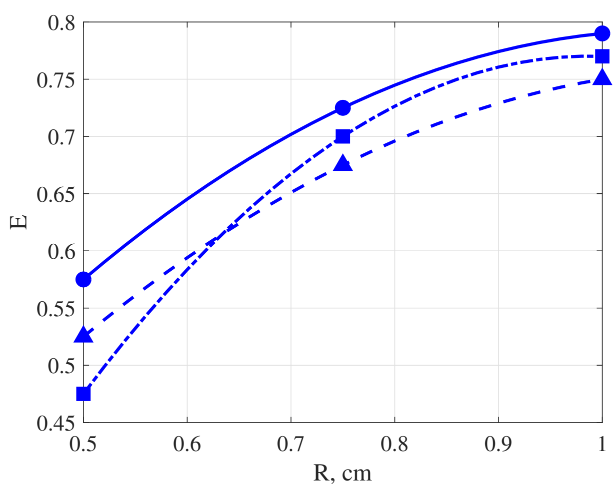

Fig. 5. Dependence of effectiveness diagnostic method according to microwave radiometry on size of tumor for various machine learning methods (SVM is circles, KNN is triangles and NBC is squares).

To determine dependence of the effectiveness diagnostics on size of the tumor, we used binary classification "Healthy" and "Cancer". This is due to the fact that most important, in our opinion, is correct detection of malignant neoplasms.

A measure of the effectiveness of medical diagnostics is considered to be the geometric mean of sensitivity and specificity

$$G = \sqrt{L \cdot S}\,, \tag{3}$$

where $L = \dfrac{TP}{TP + FN}$, $S = \dfrac{TN}{TN + FP}$, TP is proportion of correctly classified glands class "Cancer", FN is proportion of misclassified g lands c lass " Cancer", TN i s p roportion of correctly classified g lands c lass "Healthy", FN i s proportion of misclassified glands class "Healthy".

TABLE II. Sensitivity, specificity, and effectiveness of the SVM method for various data samples

	L	S	G
R=0.5 cm	0.68	0.49	0.577
R=0.75 cm	0.78	0.725	0.75
R=1 cm	0.82	0.76	0.79

The calculated sensitivity and specificity indicators (Table 2) are quite high. At the same time, we should note characteristic space limited only by temperatures and rather small training data set.

TABLE III. Results of neural network testing for various parameters

	Model 1	Model 2	Model 3	Model 4
Number of layers	6	5	4	5
Number of neurons in 1 layer	18	18	18	18
Number of neurons in 2 layer	18	18	18	9
Number of neurons in 3 layer	18	9	3	3
Number of neurons in 4 layer	18	3	2	3
Number of neurons in 5 layer	18	2	–	2
Number of neurons in 6 layer	2	–	–	–
L	0.72	0.62	0.59	0.81
S	0.66	0.59	0.55	0.67
E	0.7	0.67	0.57	0.75

Neural network testing results show comparable results with machine learning methods (Table 3). Structure of neural network significantly affects the accuracy of diagnostics. Effectiveness of diagnostics increases with increase in radius of tumor (Fig. 5.). Even for small tumors of radius R=0.5 cm with a probability of 57.5%, it is possible to correctly determine the class. We can expect a successful application of microwave radiometry method for smaller tumors, with an expansion of feature space, an increase in sample size, and the use of heuristics.

Acknowledgment

MP thanks RFBR (project number 19-37-90142) for the financial support. AK acknowledges the Ministry of Science and Higher Education of the Russian Federation (government task, project No.0633-2020-0003) for the financial support of the development of the software. EE is grateful to RFBR and the government of Volgograd region according to the research project No. 19-47-343008 for the financial support.

REFERENCES

[1] A.G. Gudkov, V.Yu. Leushin, I.A. Sidorov, S.G. Vesnin, I.O. Porokhov, M.K. Sedankin, S.V. Agasieva, S.V. Chizhikov, E.N. Gorlacheva, M.I. Lazarenko and V.D. Shashurin, "The use of multichannel microwave radiometry for the functional diagnosis of the brain," Medical Technology, vol. 2, no. 314, pp. 22-25, 2019.

[2] A.G. Losev, E.A. Mazepa and T.V. Zamechnik, "About several typical traits in the diagnosis of mammary glands pathologyaccrding to the date of microwave radiometry," Modern problems of science and education, vol. 6, pp. 254-261, 2014.

[3] A.H. Barrett and P.C. Myers, "Subcutaneous Temperature: A method of Noninvasive Sensing," Science, vol. 190, pp. 669-671, 1975.

[4] E.Y.K. Ng, "A review of thermography as promising non-invasive detection modality for breast tumor," Int J Therm Sci, vol. 48, no. 5, pp. 849-859, 2009.

[5] J.R. Keyserlingk, P.D. Ahlgren, E. Yu, N. Belliveau and M. Yassa, "Functional infrared imaging of the breast, "IEEE Eng Med Biol Mag, vol. 19, no. 3, pp. 30-41, 2000.

[6] M. Gautherie, "Temperature and Blood Flow Patterns in Breast Cancer During Natural Evolution and Following Radiotherapy," Biomedical Thermology, pp. 21-64, 1982.

[7] M.K. Sedankin, V.Yu. Leushin., A.G. Gudkov, S.G. Vesnin, D.A. Khromov, I.O. Porokhov, I.A. Sidorov, S.V. Agasieva and E.N. Gorlacheva, "Modeling the intrinsic thermal radiation of the kidney in the microwave range," Medical Technology, vol. 1, no. 313, pp. 44-47, 2019.

[8] M.K. Sedankin, V.Yu. Leushin, A.G. Gudkov, S.G. Vesnin, I.A. Sidorov, S.V. Agasieva, L.M. Ovchinnikov and N.A. Vetrova, "Applicator antennas for medical microwave radiothermographs," Medical Technology, vol. 4, no. 310, pp. 13-15, 2018.

[9] M.V. Polyakov, A.V. Khoperskov and T.V. Zamechnic, "Numerical Modeling of the Internal Temperature in the Mammary Gland," Lecture Notes in Computer Science, vol. 10594, pp. 128-135, 2017.

[10] V. Levshinskii, M. Polyakov, A. Losev and A. Khoperskov, "Verification and Validation of Computer Models for Diagnosing Breast Cancer Based on Machine Learning for Medical Data Analysis," Communications in Computer and Information Science, vol. 1084, pp. 447-460, 2019.