DOI 10.1007/s10527-022-10153-y
Biomedical Engineering, Vol. 55, No. 6, March, 2022, pp. 442-447. Translated from Meditsinskaya Tekhnika, Vol. 55, No. 6, Nov.-Dec., 2021, pp. 52-55.
Original article submitted September 28, 2021.

Some Methods for Substantiating Diagnostic Decisions Made Using Machine Learning Algorithms

A. G. Losev[1], I. E. Popov[1], A. Yu. Petrenko[1], A. G. Gudkov[2], S. G. Vesnin[3], and S. V. Chizhikov[2,4]*

Various classification algorithms used in the diagnosis of breast cancer based on microwave radiometry data are considered. In particular, their principles of operation and the possibility of substantiating diagnoses using numerical data are discussed. A substantiation algorithm based on decision trees and a naïve Bayesian classifier is presented. Examples of substantiation are given for breast cancer.

Introduction

Machine learning algorithms are currently used in medicine as a major component of consultation systems. They make diagnostic decisions with quite high levels of accuracy, significantly facilitating thereby the physician's work. However, some machine learning algorithms are able to solve another no less important task: substantiating diagnoses in a language understood by specialists. This paper considers the question of substantiation of diagnoses made using microwave radiometry.

Microwave radiometry is a biophysical method for noninvasive investigations based on measurements of internal and surface tissue temperatures from heat emission in the microwave and infrared ranges [1, 2]. Recent years have seen results in this area demonstrating the effectiveness of this method in the diagnosis of a number of diseases [2-10]. In particular, a number of studies have proposed the following approach to solving this problem [3-6]. Descriptive mathematical models characterizing the distinguishing features of the temperature fields of patients of different diagnostic classes are developed to support diagnostic decision making. The resulting mathematical models define a corresponding feature space, which is used for classification. The feature space has two functions: first, algorithms trained using these features display high diagnostic accuracy; second, as the features characterize the distinguishing properties of a disease, it becomes possible to analyze and interpret them in a language understood by diagnosticians.

The problem of substantiation in machine learning is far from new. However, it is ever more relevant. In particular, means of interpreting classifiers using text data [1] and images [11] have been considered by researchers. A variety of general approaches have been discussed [12]. However, processing of quantitative features has its specific characteristics. While in processing words and images it is sufficient to highlight points typical of one class or another, numerical data are more naturally analyzed using mathematical models, as will be demonstrated below. It is therefore appropriate to consider substantiation using quantitative data separately.

Substantiation methods can be divided into two types. Methods of the first type do not use any information from the classifier other than the diagnosis proposed. Substantiation in this case is built on the values of features and their processing by statistical methods. However, this substantiation often fails to correlate with diagnostic algorithms. Methods of the second type use classifier data with subsequent processing. For example, probabilistic information from a naïve Bayesian classifier can be used. In this case, substantiation corresponds to the principles of operation of the classifier, and the specialist can see all the grounds for making the diagnosis. In this work, only algorithms of the second type are considered.

Despite the fact that substantiation is discussed in the context of microwave radiometry, the methods proposed can also be used for diagnosis using other data. The

[1] Volgograd State University, Volgograd, Russia.
[2] Bauman Moscow State Technical University, Moscow, Russia; E-mail: chigikov95@mail.ru
[3] RES Company, Moscow, Russia.
[4] OOO NPP Tekhnologicheskie Innovatsii, Moscow, Russia.
* To whom correspondence should be addressed.

key point in substantiation is a feature space suitable for interpretation in a language understood by specialists.

Materials and Methods

Diagnosis of breast cancer was made based on measurements of skin and internal temperatures at 10 different points in each breast, along with two further reference points. The measurement scheme is shown in Fig. 1. In this scheme, the central point corresponds to the nipple and points 1-8 to the external radius of the breast. T1 and T2 are reference points.

Thus, the database consisted of a set of internal temperatures (1) and a set of skin temperatures (2):

$$T^{i,mw} = \left\{t^{mw}_{0,r}, ..., t^{mw}_{9,r}, t^{mw}_{0,l}, ..., t^{mw}_{9,l}, t^{mw}_{0,p}, t^{mw}_{1,p}\right\}; \quad (1)$$

$$T^{i,ir} = \left\{t^{ir}_{0,r}, ..., t^{ir}_{9,r}, t^{ir}_{0,l}, ..., t^{ir}_{9,l}, t^{ir}_{0,p}, t^{ir}_{1,p}\right\}, \quad (2)$$

where i is the patient number, mw indicates temperatures measured in the microwave range, and ir indicates temperatures measured in the infrared range. The first subscript is the number of the point at which the temperature was measured; the second subscript is the measurement location: l – left breast, r – right breast, p – reference area.

The following characteristic features were identified in the process of analyzing temperature measurements as typical of healthy subjects and generally not of patients without breast cancer: low levels of thermal asymmetry (the difference between temperatures at symmetrical measurement points in right and left breasts); uniform distribution of temperature in the breast, i.e., absence of "hot spots"; particular values for the skin-to-internal temperature ratio.

These data were used to construct 62 different features [5]. The resulting feature space was classified using various algorithms. Table 1 shows the results of computational experiments. Algorithms were evaluated using two parameters: sensitivity (the proportion of correctly diagnosed patients in the risk group) and specificity (the proportion of correctly diagnosed healthy subjects). This yielded a feature space effective for the diagnosis of breast cancer and providing substantiation in a language understood by specialists.

TABLE 1. Classifier Accuracy

Classifier	Sensitivity	Specificity
Neural network	0.930	0.908
Logistical regression	0.894	0.913
Decision tree	0.886	0.889
Naïve Bayesian classifier	0.860	0.867

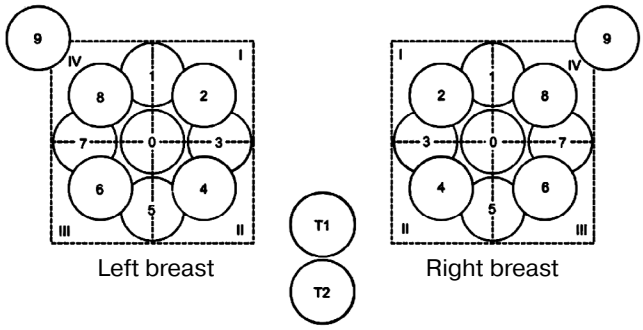

Fig. 1. Temperature measurement scheme.

Let us consider these algorithms one by one.

Neural Networks

This model consists of a set of artificial neurons combined into layers. Neurons from one layer send their information to neurons in the next layer. Each neuron is a combination of neurons in the preceding layers. The first layer consists of neurons whose values correspond to the input feature. This type of classifier is often highly accurate, though it has one major drawback – uninterpretability. Because of the constant transformation of features, the model becomes a "black box," in which it is not possible to evaluate the contributions of each feature to the diagnosis and the corresponding diagnostic grounds. It is important to note that this applies to numerical data. As already noted in the Introduction, these have their specific characteristics. In particular, neural networks operating with images are interpretable.

Logistical Regression

This algorithm is a probabilistic model. Diagnoses are established in accordance with the logistical function

$$f(z) = \frac{1}{1+e^{-z}}, \quad (3)$$

where

$$z = \sum_{i=1}^{n} c_i \cdot x_i$$

is a linear combination of features, x_i are features, and c_i are regression coefficients. If the value of the linear combination is negative, the subject is regarded as healthy. Conversely, a positive value is indicative of a disease. Work reported in [13] describes the process of substantiation of diagnoses by logistical regression. However, we believe that it is important to note some drawbacks and limitations of the algorithm.

The principle of operation of the algorithm is such that the substantiation is based on the values of products c_ix_i. If the product is negative, feature x_i is typical of healthy subjects. However, in practice, features often have exclusively positive or negative values. That is, $\text{sign}(c_ix_i) = \text{const}$, so that the feature becomes typical of only one diagnosis. To solve this problem, the feature needs to be transformed such that a negative value stays for one diagnosis and a positive value, for another. Such transformation of features requires a separate study and is not addressed here.

Decision Tree

This algorithm is a conditional model. The conditions are organized into a tree-like structure whose nodes represent conditions of the type "Feature $i \le x$." Each node is connected with the next two. The final node is the diagnosis. Figure 2 presents an example of a decision tree.

The key point in substantiation is that the structure of the decision tree is static, so each node can be given a verbal description. For example, in Fig. 2, the first node might correspond to the following descriptions:

– if the condition at the node is fulfilled, then "The value of the feature 1 is normal;"

– if the condition at the node is not fulfilled, then "The value of the feature is 1 is significantly elevated."

The principle of assessing the "significant elevation" is analogous to the applications of the naïve Bayesian classifier considered below. However, the nodes of the tree are described manually, as the values of each subsequent node depend on the context of the previous node.

1. **For each** node **in** the decision tree:

2. **If** the patient being classified "falls" into the node under consideration, **then:**

3. **If** the condition at the node is fulfilled for the patient, **then**: display the information for the node when the condition is fulfilled.

4. **Else:** display information for the node, when the condition is not fulfilled.

Naïve Bayesian Classifier

This algorithm is a probabilistic model that makes diagnoses on the basis of the object's features being typical of one or another class of patients. In our opinion, it is convenient for substantiation purposes. The formula for computation of the class is:

$$\hat{y} = \arg\max_y P(y)\prod\nolimits_{i=1}^{n} P(x_i \mid y), \quad (4)$$

where $\hat{y}$ is the final class; $P(y)$ is the probability of class y, i.e., the proportion of the class in the training set; and $P(x_i \mid y)$ is the conditional probability.

Thus, if $P(x_2 \mid \text{sick}) > P(x_2 \mid \text{healthy})$, this means that feature x_2 is characteristic of subjects of the "sick" class. The substantiation is based on this knowledge.

The substantiation algorithm in more details is:

1. **For each** feature i **of** [feature 0 to feature m]:

2. **If** $P(\text{feature } i \mid 1) > P(\text{feature } i \mid 2)$, **then**: Feature i is characteristic of class 1 (healthy subjects).

3. **Else**: Feature i is characteristic of class 2 (patients in the risk group).

4. **Determine** the deviation of feature i from normal.

As probabilistic evaluation of the classifier provides no information as to what is specific about the feature for a given class, additional analysis of features is required. One approach to this analysis consists of comparing the value of the feature with its distribution in healthy subjects. Let us compare the feature with four percentiles (25th, 40th, 60th, and 75th):

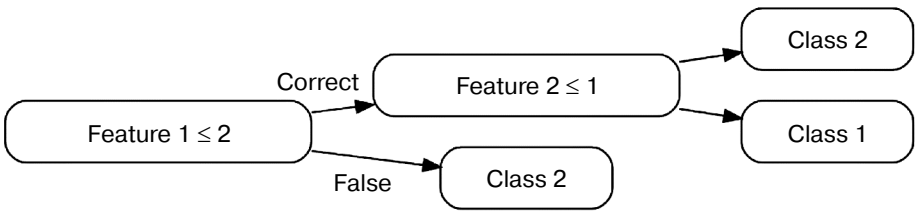

Fig. 2. Example of decision tree structure.

– if below the 25th percentile, it is significantly below normal;

– if between the 25th and 40th percentiles, it is slightly below normal;

– if between the 40th and 60th percentiles, it is within the normal range;

– if between the 60th and 75th percentiles, it is slightly greater than normal;

– if greater than the 75th percentile, it is significantly greater than normal.

Thus, substantiation of the type "Mean temperature of the left breast is significantly greater than normal, which is characteristic of patients with cancer" is obtained. The number of threshold values and their description can be altered depending on the specific details and context of the subject area.

Results

We will present an example of substantiation using the "naïve Bayesian classifier" and "decision tree" algorithms. The feature space will consist of four functions characterizing a number of hypotheses:

– insignificant thermal asymmetry (feature F1):

$$\left\| T_r^{i,mw} - T_l^{i,mw} \right\|_1 = \sum_{k=0}^{9} \left| T_r^{i,mw} - T_l^{i,mw} \right|;$$

– uniform temperature distribution (feature F2):

$$\max_{t \in T^{i,mw} \setminus T_0^{i,mw}} \left| T_0^{i,mw} - t \right|;$$

– internal temperature variation. Increased values indicate the presence of a "hot spot" (feature F3):

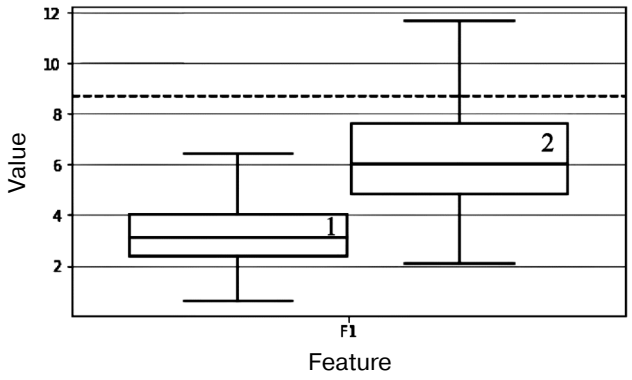

Fig. 3. Box and whisker plot for the first feature: 1) healthy; 2) sick.

TABLE 2. Naïve Bayesian Classifier

Feature	F1	F2	F3	F4
Value	8.7	1.5	3	0.89
P(Feature | class "healthy")	$4.3 \cdot 10^{-7}$	0.26	0.1	0.18
P(Feature | class "sick")	0.1	0.42	0.17	0.29

$$\max_{t \in T^{i,mw}} t - \min_{t \in T^{i,mw}} t;$$

– skin temperature variation. Increased values indicate the presence of a "hot spot" (feature F4):

$$\max_{t \in T^{i,jr}} t - \min_{t \in T^{i,jr}} t.$$

Let us first consider the naïve Bayesian classifier. The values of the features for a subject undergoing diagnosis and their conditional probabilities are given in Table 2.

It can be seen from Table 2 that the probability of the subject belonging to the class of patients is evidently greater in terms of all features than the probability of belonging to the group of healthy subjects. We will consider this using the first feature as an example. Figure 3 shows a box and whisker plot. The box at left shows the distribution of values for healthy subjects and the box at right shows values for patients. The dotted line shows the value of the feature in the subject concerned. It can be seen from Fig. 3 that such values for features are encountered more frequently in patients. This also applies to other features.

Thus, we obtain the following substantiation according to this algorithm:

– all features are typical of the "sick" class;

– thermal asymmetry values are significantly greater than normal;

– nipple temperature is slightly greater than normal;

– internal temperature variation is significantly greater than normal; internal "hot spots" may be present;

– skin temperature variation is slightly greater than normal.

Let us consider a decision tree using the same example. Using the training set, the algorithm constructed a tree structure shown in Fig. 4.

On diagnosis, the subject checks conditions from left to right. Thus, grey cells in Fig. 4 are those complying with the corresponding rules for the subject. In the case under consideration, increased thermal asymmetry indicates that the subject is sick. Feature F3 – the magnitude of variation in deep temperature – is not included in the substantiation. This is because the value of 3.85 for this feature is significantly greater than normal. The fact that

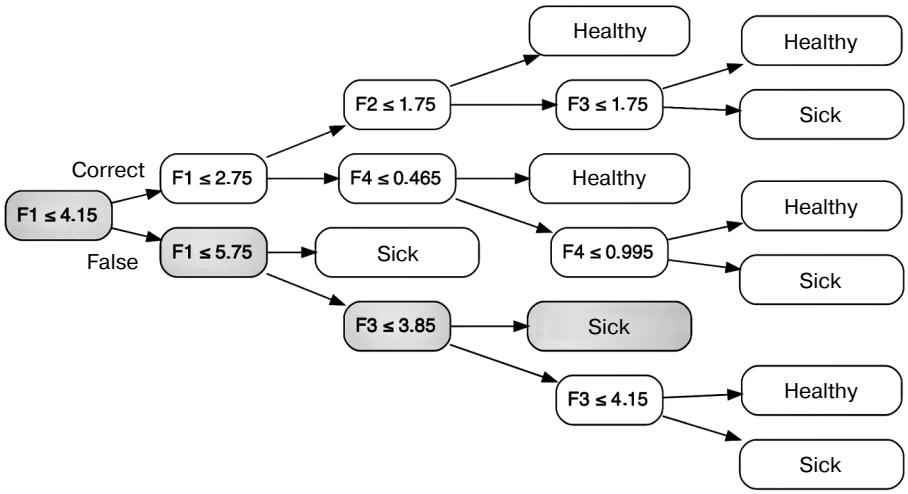

Fig. 4. Decision tree structure.

the value of a feature is below a specified level does not characterize it. It could be either at the normal level or significantly different from normal. However, this piece of information is not used in the diagnosis. It is sufficient for the classifier that the subject has elevated thermal asymmetry.

Let us consider a healthy subject correctly identified by both classifiers. According to the naïve Bayesian classifier, the values of each feature are more characteristic of healthy subjects. Comparison of features with normal gave the following results: thermal asymmetry is in the normal range, nipple temperature is slightly below normal, internal temperature variation is slightly below normal, and skin temperature variation is significantly below normal.

Substantiation using a decision tree yielded the following: thermal asymmetry is not increased, and nipple temperature is low.

It is important to note that substantiation using different algorithms does not guarantee obtaining similar results, as seen in the examples given.

Conclusions

It should be noted in conclusion that selection of a substantiation method depends primarily on the diagnostic accuracy of the classifier. If diagnostic accuracies of all classifiers differ insignificantly, the choice is based on the specific characteristics of the substantiation methods. In one area it might be important to report probabilistic information, while in another it may be more important to have relative statistical information, i.e., increases or decreases in feature values relative to normal.

In our opinion, further studies should address other algorithms and the development of methods for substantiating these algorithms using numerical data. Based on the classification algorithms considered here, we see no opportunity for building a general approach to substantiation, as substantiation is significantly influenced by the specific characteristics of operation of each algorithm.

This work was funded by the Russian Science Foundation (Project No. 19-19-00349 for the hardware part) and the Foundation for Assistance to Small Innovative Enterprises in Science and Technology (Project No. 136GRTsTS10-D5/61904 for the artificial intelligence part).

REFERENCES

1. Ribeiro, M. T., Singh, S., and Guestring, C., "'Why should I trust you?': Explaining the predictions of any classifier," in: Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 13-17, 2016, San Francisco, California, USA, pp. 1135-1144.
2. Vesnin, S. G., Turnbull, A. K., Dixon, J. M., and Goryanin, I., "Modern microwave thermometry for breast cancer," J. Mol. Imag. Dynam., **7**, No. 2, 1-6 (2017).
3. Zamechnik, T. V., Losev, A. G., and Levshinskii, V. V., "Results of optimization of diagnostic features of breast cancer detected by microwave radiometry," Med. Vestn. Severn. Kavkaza, **14**, No. 1.1, 48-52 (2019).

4. Levshinskii, V. V., "Intelligent system for diagnostics of venous diseases based on the microwave radiometry data," Lecture Notes in Networks and Systems, **155**, 212-219 (2021).
5. Levshinskii, V. V., "Mathematical models for analyzing and interpreting microwave radiometry data in medical diagnosis," J. Comput. Eng. Math., **8**, No. 1, 3-14 (2021).
6. Levshinskii, V., Galazis, C., Ovchinnikov, L., Vesnin, S., Losev, A., and Goryanin, I., "Application of data mining and machine learning in microwave radiometry (MWR)," Commun. Comp. Inform. Sci., **1211 CCIS**, 265-288 (2020).
7. Osmonov, B., Ovchinnikov, L., Galazis, C., Emilov, B., Karaibragimov, M., Seitov, M., Vesnin, S., Losev, A., Levshinskii, V., Popov, I., Mustafin, C., Kasymbekov, T., and Goryanin, I., "Passive microwave radiometry for the diagnosis of coronavirus disease 2019 lung complications in Kyrgyzstan," Diagnostics (Basel), **11**, No. 2, 1-15 (2021).
8. Goryanin, I., Karbainov, S., Shevelev, O., Tarakanov, A., Redpath, K., Vesnin, S., and Ivanov, Y., "Passive microwave radiometry in biomedical studies," Drug Discov. Today, **25**, No. 4, 757-763 (2020).
9. Tarakanov, A. V., Tarakanov, A. A., Vesnin, S. G., Efremov, V. V., Roberts, N., and Goryanin, I., "Influence of ambient temperature on recording of skin and deep tissue temperature in region of lumbar spine," Eur. J. Molec. Clin. Med., **7**, No. 1, 21-26 (2020).
10. Gudkov, A. G., Leushin, V. Y., Vesnin, S. G., Sidorov, I. A., Sedankin, M. K., Solov'ev, Y. V., Agasieva, S. V., Chizhikov, S. V., Gorbachev, D. A., and Vidyakin, S. I., "Studies of a microwave radiometer based on integrated circuits," Biomed. Eng., **53**, No. 6, 413-416 (2020).
11. Spinks, G. and Moens, M.-F., "Justifying diagnosis decisions by deep neural networks," J. Biomed. Informat., **96**, 1-13 (2019).
12. Loi, M., Ferrario, A., and Viganò, E., "Transparency as design publicity: Explaining and justifying inscrutable algorithms," Ethics Inf. Technol., 1-20 (2020).
13. Biran, O. and McKeown, K., "Human-centric justification of machine learning predictions," in: Proc. 26th International Joint Conference on Artificial Intelligence, August 19-25, 2017, Melbourne, Australia, pp. 1461-1467.