Using Neural Network Algorithms in Prediction of Mean Glandular Dose Based on the Measurable Parameters in Mammography

DENIS ČEKE1, SUAD KUNOSIC2,3, MUSTAFA KOPRić2,3, LIDIJA LINCENDER4

University Center for Distant Education Development, University of Tuzla, BiH1
Department of Physics, Faculty of Natural Sciences and Mathematics, University of Tuzla, BiH2
Department of Biophysics, Medical Faculty, University of Tuzla, BiH3
Faculty of Health Studies, University of Sarajevo, BiH4

SUMMARY. In this paper we were investigate possibility of using neural network algorithms in prediction of mean glandular dose (MGD), based on the measurement of the compressed breast thickness (CBT) in patients population between 40 – 65 years. According to the available information this is the first time that is someone explored this possibility of using neural networks in prediction of MGD based on the information of CBT. The primary aim of this method is reducing unnecessary overdose of X-ray exposure to patients. The study population consisted of 63 patients (234 screens) from 40 to 64 year during routine mammographic control. The best results were achieved with Levenberg-Marquardt learning algorithm where correlation factor between neural network outputs and targets was $R=0.845$ ($71.4\%$).

Keywords: neural network, mammography, mean glandular dose (MGD), compressed breast thickness (CBT).

1. INTRODUCTION

An objective of mammography screening is early detection of breast cancer. At present, mammography is the most accurate and reliable means of detecting minimal, non-palpable breast cancer [1]. Carcinogenic risk associated with mammography and the absorbed radiation applied to the breast has given rise to concern. Relationship between MGD and CBT is commonly used for the presentation of mammographic dose survey results and could be useful for the assessment of individual breast doses retrospectively for situations where dose measurements cannot be provided continuously in house [2]. Theoretically, potential risk of radiation influence on patient’s health during this procedure is very small in comparison with benefit that is achieved during regular diagnostic procedure in detection of possible malignant lesions [3, 4, 5, 6]. Calculating MGD based on the measurable parameters in mammographic procedure is time consuming task and it is very hard to do calculation in the time of patient processing. In any radiographic procedure, it is imperative that the radiation dose is as low as reasonably practicable, while maintaining an adequate image quality. This is particularly important in radiography of sensitive organs such as breast and in screening programs where the exposed population is asymptomatic. It is generally accepted that the glandular tissue of breast is the most radiation-sensitive tissue [7]. Therefore, the suggestion that the mean glandular dose is the most appropriate dosimetric quantity to predict the risk of radiation-induced carcinogenesis has been widely accepted [7, 8].

Neural network model developed in this work were used for prediction of MGD in mammography based on CBT and other measurable parameters.

2. RESEARCH GOAL

Primary research goal of this work was to investigate a possibility of using neural networks in prediction of mean glandular dose based on the information about CBT and other measurable parameters. Direct measurement of MGD is very hard to calculate because of complicated mathematical operations that need to be done. In this work we also investigate a possibility of constructing model that will define the possible intelligent solution to this problem.

3. MATERIALS AND METHODS

The input variables to the system were collected as measurable variables during routine mammography procedure. The data about patients were collected from Department of Thoracic Diagnostic with Breast of the Radiology Clinic (of the University of Sarajevo Clinics Centre). Data about patient’s doses were collected for 63 patients (234 screens) from 40 to 64 year during routine mammographic control.

For preparing system input variables we have created MATLAB...
function that calculate non measurable variables (MGD) based on measurable ones. Measurable variables that were used as input to system were: age of patient, used clinical spectra, compressed breast thickness and type of projection, exposition factors and charge (mAs), anode voltage (kVp). Non measurable variable that were used as input to the system were MGD. Patient’s doses were calculated according to the recorded data.

MGD for each mammogram is defined on a basis of conversion factors calculated by Dance et al (9) and a calculated ESAK (entering air kerma measured freely in air without backscatter), using the following relation [10]:

\[
\text{MGD} = g \cdot c \cdot s \cdot \text{ESAK}
\]

ESAK for each individual exposition is calculated from post – exposure mAs (I \cdot t) and output data for the x – ray set in μGy mAs⁻¹ used in an exposition field. Conversion factors were calculated by Dance, for a different clinical spectrum (target/filter combination), half value layer (HVL), compressed breast thickness and breast glandularity. Factor \( g \) defined by Dance et al [9] corresponds to glandularity of 50 % while factor \( c \) includes every change in breast glandularity of 50 %. Factor \( s \) was defined by Dance et al [9] for a normal composition and for various thickness of a compressed breast (20 – 110 mm) and HVL. Two tables of \( c \)-factors for average breast compositions have been calculated for women attending screening in the normal age range of 50 – 64 and for those attending the age trial in the age range 40 – 49 years [9]. Finally, factor \( s \) includes a correction for used type of the clinical spectra Mo/Mo.

In this research we were focused on developing of Multi Layer Perceptron (MLP) neural network model which we were trained with backpropagation algorithm available in MATABLAB environment. Before passing variables to the neural network we had to make preprocessing unit for calculating MGD. The schematic representation of created system is given in Fig. 1.

Backpropagation algorithm is used to update the weights and biases of the neural network. Backpropagation network with biases, a sigmoid transfer function in the hidden layer, and linear transfer function in the output layer is capable of approximating any function [12]. Weights and biases are updated using variety of gradient descent algorithms. The gradient is determined by propagating the computation backwards from output layer to first hidden layer. If it is properly trained, the backpropagation network is able to generalize to produce the reasonable outputs on inputs it has never “seen”, as long as the new inputs are similar to the training inputs.

The operation of a typical backpropagation network can be described as follows [13]:

1. After presenting signals to the input layer, information propagates through the network to the output layer (forward propagation). During this time input and output states for each neuron will be set.

\[
x^{(s)}_j = f(I^{(s)}_j) = f \left( \sum_{i} (w^{(s)}_{ji} \cdot x^{(s-1)}_i) \right)
\]

where

\( x^{(s)}_j \) denotes the current output state of the \( j \)-th neuron in the current \( s \)-layer,

\( I^{(s)}_j \) denotes the weighted sum of inputs to the \( j \)-th neuron in the current \( s \)-layer,

\( f \) is conventionally the sigmoid function,

\( w^{(s)}_{ji} \) denotes the connection weight between the \( i \)-th neuron in the surrounding layer \( s \) and the \( j \)-th neuron in the previous layer \( s-1 \).

2. Global error is generated based on the summed difference of required
and calculated output values of each neuron in the output layer (Eq. 3).

\[ E_{glob} = \frac{1}{2} \sum_k ((r_k - o_k)^2) \]

where

\[ E_{glob} \] means the global error,

\[ (r_k - o_k) \] denotes the difference of required and calculated output values.

Scaled local error for each neuron in the output layer is calculated according to the following formula (Eq.4.):

\[ e^{(0)}_j = x^{(0)}_j \cdot (1 - x^{(0)}_j) \cdot (r_k - o_k) \] (4)

This formula shows that local errors are scaled, based on their output activation values.

3. Global error is back-propagated through the network to calculate local error values and delta weights for each neuron (Eq. 5. and 6.). Delta weights are modified according to the Delta-rule that strictly controls the continuous decrease of the synaptic strength of those neurons that are mainly responsible for the global error. In this manner, a regular decrease of global error is assured.

\[ \Delta w^{(s)}_{ij} = lcoef \cdot e^{(s)}_i \cdot x^{(s)}_j \] (6)

where \( \Delta w^{(s)}_{ij} \) denotes delta weight of the connection between current neuron and the joining neuron, \( lcoef \) denotes the learning coefficient, one of the training parameters.

4. Synaptic weights are updated by adding delta weights (\( \Delta w^{(s)}_{ij} \)) to the current weights.

Our aim was not to build a classifier, but to design a neural network capable of giving results in the preset continuous interval, which is more of regression problem.

The major problem when designing neural network is choosing proper topology and learning algorithm. There are a lot of learning algorithms that are available in MATLAB environment that are variations of backpropagation algorithm, some of them are fast, and some of them are slow. In this work we were decided to fix the network topology and test some of the training algorithms. Our network consist of three layers, input, hidden and output layer. Transfer function in the hidden layer is tan-sigmoid. Hidden layer have five neurons. Output layer uses linear transfer function and have one output, predicted MGD.

The following learning algorithms were tested in the MATLAB environment:

- Scaled Conjugate Gradient
- Powell-Beale Restarts
- BFGS (Broyden, Fletcher, Goldfarb, Shanno)
- One Step Secant
- Levenberg-Marquardt
- Resilient Backpropagation

4. RESULTS

After the training sessions we used a regression analyze between the network response and the corresponding targets. For this purpose we used MATLAB postreg routine which returns three parameters, \( m \), \( b \) and \( R \) value [14]. The \( m \), and \( b \) correspond to the slope, and \( y \) intercept of the best linear regression relating targets (T) to network outputs (OUT). A perfect fit (output exactly equals to targets) has \( m \) and \( b \) equal to 1 and 0, respectively. The \( R \) value is measure of how well the variation in the outputs is explained by the targets. The perfect correlation between the targets and the outputs has \( R \) value equal to 1. If \( r \) value is very small or equal to 0, that means that correlation between the targets and outputs is very small or there is no correlation at all, respectively.

Fig. 3a and 3b presents correlation results for Scaled Conjugate Gradient and Powell-Beale Restarts algorithms, respectively. They both belong to conjugate gradient algorithms with fast convergence. Correlation between training outputs and targets for Scaled Conjugate Gradient and Powell-Beale Restarts algorithms were \( R=0.803 \) and \( R=0.816 \), respectively.

Fig. 4a and 4b present simulation results from two Quasi-Newton algorithms, BFGS and One Step Secant algorithm, respectively. They are also learning algorithms with...
fast convergence. In the case of these two learning algorithms, BFGS and One Step Secant algorithm, correlations were $R=0.826$ and $R=0.756$, respectively.

Fig. 5a and 5b present results of simulation for Levenberg-Marquardt and Resilient Backpropagation learning algorithms with correlation $R=0.845$ and $R=0.821$, respectively.

The coefficient of determination $R^2$ is a measure of how well the regression line represents the data. If the regression line passes exactly through every point on the scatter plot, it would be able to explain all of the variation. The further the line is away from the points, the less it is able to explain. Finally, it provides a measure of how well future outcomes are likely to be predicted by the model [15]. The coefficient of determination is such that $0 \leq R^2 \leq 1$, and denotes the strength of the linear association between input and output values. Bigger $R^2$ corresponds to better correlation between neural network output and desirable values.

Summary of results for all training algorithms is shown in Table 1.

After the neural network model has been tested with presented learning algorithms it is shown that Levenberg-Marquardt algorithm have the best performance. Correlation factor, $R$ value for this algorithm is 0.845, coefficient of determination, $R^2$, have value of 0.714 and percent of common variance (PCV=$R^2\times100$) is 71.4%.

Results for BFGS and Resilient backpropagation algorithm were the closest to Levenberg-Marquardt’s algorithm and One Step Secant had the worse result.

5. CONCLUSION

In this work we have described the scheme of developed MLP neural network model, which had been trained with different training algorithms to predict MGD based on the measurable parameters. Test on the database of values that we have collected gave very promising results. In all cases correlation factor $R$ was above 0.80 except for One Step Secant algorithm which gave us correlation factor value of 0.75.

There has been very little published literature about this kind of approach in predicting of MGD. Although in principle, this is not standard way of defining MGD in routine mammography, the current method of employing neural networks in predicting MGD has the advantage in knowing possible MGD of patient before one is exposed to $x$-rays which can have possible bad consequences on the current health of the patient. This is the first step in developing such neural network system which can be trained as more data is available to produce even better results.

**REFERENCES**


**Corresponding author:** Denis Ceke, MSc.
University Center for Distant Education Development. University of Tuzla, BiH. e-mail: denis.ceke@untz.ba