Volume 20, No 3 | International Journal of Radiation Research, July 2022

Optimizing the neural network training algorithm in predicting kerma in mammography

M. Nabipour[1], M.R. Deevband[1*], A. Asgharzadeh Alvar[1], N. Soleimani[2]

[1]Biomedical Engineering and Medical Physics Department, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[2]School of Medicine, Golestan University of Medical Sciences, Gorgan, Iran

► ***Original article***

***Corresponding author:**
Mohammad R. Deevband, Ph.D.,
E-mail:
mdeevband@sbmu.ac.ir

Received: May 2021
Final revised: February 2022
Accepted: February 2022

Int. J. Radiat. Res., July 2022; 20(3): 665-670

DOI: 10.52547/ijrr.20.3.21

Keywords: *Breast mammography screening, entrance surface air kerma estimation, MLP neural network, RBF neural network.*

ABSTRACT

Background: In regard to the enhanced use of mammography screening tests for screening breast cancer, some concerns on the enhancement of the patient's absorbed dose have increased as well. Therefore, the assessment of the patient's dose before mammography is very important, and being aware of the dose level by its estimation can be helpful before radiation. ***Materials and Methods***: To this end, an artificial neural network (ANN) was used in this study to estimate the entrance surface air kerma (ESAK). A phantom with similar characteristics of the breast tissue was also used to collect the required data and the network was trained using some measurable parameters. To conduct the current research, multilayer perceptron (MLP) neural network architecture with training algorithms of LMBP, SCGBP, Rprop, BFGS, and GDBP, as well as radial basis function (RBF) neural network were used. ***Results***: The results show that the neural network with BFGS training algorithm and 38 hidden layer neurons has the best performance with 7.40% root mean square error (RMSE) and coefficient of determination (R^2) was obtained as 0.91. ***Conclusion***: According to the results of this study, there is a good correlation between the estimated network output and the measured values of the ESAK. The present method will remove the limitations and costs associated with the preparation process of dosimeter instruments.

INTRODUCTION

According to the reports of the National Cancer Institute, following lung cancer, breast cancer is known as the second deadliest cancer, and about 43,000 women annually die due to breast cancer in the United States [1]. The World Health Organization has also reported that about 5000 women died due to breast cancer in Iran in 2020 [2].

Mammography is known as the most effective diagnostic method for breast cancer at an early stage; therefore, it is used for early detection of breast cancer [3]. One of the limitations of mammography is the low contrast between pathological and normal tissues. This limitation makes it possible to use low-energy photons in mammography, which can be easily absorbed in the breast tissue and then increase the received dose by the person under the test [4]. However, in mammography screening, asymptomatic women are exposed to radiation [5]; therefore, in each case, the radiation dose must be kept at the lowest possible level while maintaining suitable image quality. In addition, there is always an accidental risk of developing breast cancer in mammography due to exposure to breast tissue. In this regard, the evaluation of the patient's dose during mammography is important [6]. It is generally accepted that glandular tissue is the most sensitive tissue in the breast to radiation. Thus, the mean glandular dose (MGD) is suggested as the most appropriate dosimetry, in order to predict the risk of radiation-induced cancer [7,8], which is necessary to achieve ESAK.

One of the common ways used to calculate the MGD is Monte Carlo simulation. Using this method, the mammography machine is simulated with its details, which requires performing a separate simulation for each center and device due to the differences in device characteristics, which is time-consuming and complex [9]. Moreover, there is an error due to the differences in the function of devices with their nominal characteristics, depending on the lifetime of each device [9,10]. In a study conducted by Suleiman using Bland-Altman analysis and regression to investigate compatibility and correlation between organ dose and calculated dose, a significant bias was observed between these two doses [11]. In another study, the mean MGD error was calculated as 3.80% by Dance parameters compared to the dose measured by dosimeter [5].

Using another method, it is needed to measure ESAK of breast tissue and use the conversion factors

[DOI: 10.52547/ijrr.20.3.21]

calculated by Dance, which would be time-consuming and difficult when processing patients. Accordingly, in this method, the breast tissue dosimetry result is also required, but due to the limitations in dosimetry instruments, measurement-based methods cannot be efficient enough for estimating the dose (12,13).

The neural network has a special ability in finding nonlinear relationships between inputs and outputs variables. As well, it does not require any specific function to express the relationship between input and output data. Nowadays, researchers have shown that the use of machine-learning techniques with their diagnostic function is better than the use of linear regression and logistics that can be effective on having a faster and low-cost diagnosis in all treatment sections (14). In a study performed by Massera, deep learning was proposed to estimate volume glandular fraction based on mammographic images, which were converted to glandularity values for MGD calculations. In this study, MLP neural network architecture was also used and network prediction by R^2 was compared with the underlying criterion (15).

The present study aimed to use the optimal neural network to estimate ESAK for predicting MGD, and to provide a pre-imaging dose prediction as a necessity. It is hoped that the proposed model eliminates the error between the absorption dose measured by the dosimeters and calculations using the Monte Carlo simulation method.

MATERIALS AND METHODS

Data collection

The humanlike phantom with the trade name of Pro-DigiMAM Mark II (SN: M-DT-000152: POLAND) with breast tissue characteristics (including glandularity of 50% and a thickness of 45 mm) that complies with the international standard IEC 61223-3-2 was used for the production and collection of data. In the present study, for ESAK and half-value layer (HVL) measurement, a solid state dosimeter (RTI Electronics AB- Barracuda Cabinet BC1-11020059) was used, so the dosimeter was placed under the compressor plate of the mammography unit, next to the phantom. Of note, the ESAK and HVL values of 224 phantom dosimetry samples used in this research were collected from 32 mammography centers in Iran and some exposure parameters such as anode voltage, tube current, and anode/filter combinations were then recorded.

ESAK definition

According to Dance study (16), the MGD is defined as equation (1).

$$MGD=ESAK.g.c.s \quad (1)$$

Where ESAK is the incident air kerma at the upper surface of the breast, which is measured without any backscatter. Accordingly, the conversion factors were calculated for various clinical spectra such as HVL, anode/filter combinations, the compressed breast thickness, and the rate of breast glandularity. As well, g stands for the incident air kerma to MGD conversion factor, which is calculated by Dance and corresponds to a glandularity rate of 50%. Notably, Factor c is the correct coefficient of g for any difference in breast composition from 50% glandularity. The c-factors calculated by Dance were defined for glandularity rate between 0.1 and 100% for breast thickness of 2-11 cm and HVLs of 0.30-0.60 mm Al. Finally, the factor s is the correct coefficient of factor g for the other x-rays spectrum, which changes in the anode/filter combinations (16).

Data processing

In the current work, MATLAB software was used for processing the obtained data and modeling a neural network. Afterward, the network was trained using data prepared from phantom dosimetry in mammography centers. The measurable variables, including tube current (mAs), anode voltage (Kvp), the thickness of HVL, the overall thickness of the filter, and anode/filter combinations were applied as inputs to the network, and ESAK was applied as the network output. Before the transmission of these variables to the neural network, it is necessary to pre-process the data in order to facilitate the procedure of neural network training. Pre-processing the data includes some stages as encoding discrete inputs, management of outlier data, and data normalization. For this purpose, 11 outlier data were discarded and data normalization was then performed based on equation (2) which led the data to be placed between zero and one:

$$x_{normal} = \frac{x_{max} - x_i}{x_{max} - x_{min}} \quad (2)$$

Where x_i is an original value, and x_{normal} is the normalized value (17). Thereafter, these data were divided into three training data sets, consisting of 70% of the total data, and in this regard, each validation and testing set consists of 15% of the total data. The schematic of neural network design for the prediction of ESAK is shown in figure 1.

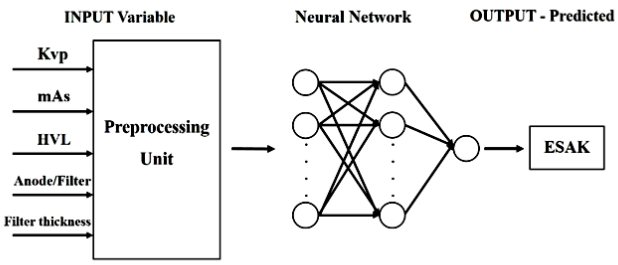

Figure 1. Schematic of neural network design for the ESAK prediction with 5 inputs.

[DOI: 10.52547/ijrr.20.3.21]

ANN architecture

MLP neural network with a sigmoid transfer function in the hidden layer as well as a linear transfer function in the output layer is known as a comprehensive estimator [18] shown in figure 2. After the selection of the base structure of the network, it is necessary to determine the details of the MLP network architecture such as the number of layers and neurons, and the training algorithm. The standard process of selecting the number of layers is performed by beginning the training process with a network consisting of one hidden layer. It is noteworthy that the use of more than two hidden layers in artificial neural networks is unconventional [17]. The number of neurons in the output layer is equal to the target data vector, in such a way that our network with only one output (ESAK) had only one output layer. The standard process of determining the number of neurons is to begin the process of training neural network with a greater number of neurons, and after several stages of training and testing the network with some hidden neurons, the number of optimal neurons is finally selected based on the performance index [19].

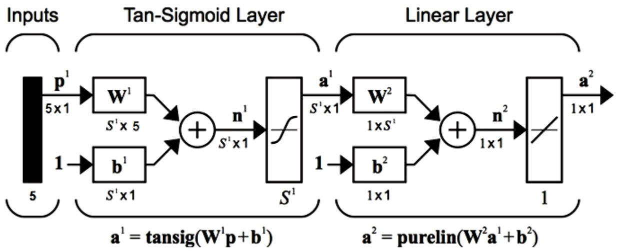

Figure 2. Two-layer neural network with 5 inputs and the tan-sigmoid transfer function and one output with the linear transfer function.

ANN training algorithms

In regard to training the neural network, the weighting coefficients of the network changed in order to minimize the performance function like the mean squared error (MSE). In this research, the most important learning algorithms such as gradient descent backpropagation (GDBP), BFGS quasi-Newton method, resilient backpropagation (Rprop), scaled conjugate gradient backpropagation (SCGBP), and Levenberg-Marquardt backpropagation (LMBP), based on backpropagation algorithm in the MLP network, which are the most widely used algorithms in function approximation problems, have been discussed [17]. Additionally, RBF networks are widely used for the nonparametric estimation of multidimensional functions [20]. Correspondingly, they have two layers, the hidden layer of which is a radial basis (such as Gaussian) function and the output layer is linear [14,18]. Adjustable important parameters include the following three parameters: spread, which shows the variance of radial basis functions; goal, which is the MSE; and the maximum number of neural network neurons with the same number as radial basis functions [21].

Data validation

After the neural network training, it is necessary to demonstrate its efficiency. The performance of the models is mostly evaluated by a set of test data that are not used in the training of the network. In the present study, the applied indicators in the evaluation of models were RMSE and R^2 between the predicted and measured values. Moreover, RMSE value measures both the accuracy and validity of training and test data sets. The R^2, which is the square of the correlation coefficient, is one of the most widely used statistics in the articles related to the neural network. In addition, the histogram of network errors was used to evaluate the network [18].

RESULTS

In the current research, by executing the program for the number of neurons from different hidden layers, the optimum number for GDBP was determined to be between 5 and 11 neurons. However, for other training algorithms, it was between 32 and 38 neurons, and in higher numbers, regression was lower than 80% due to the complexity of problem-solving and calculations. Tables 1 and 2 show the values of RMSE and R^2 of neural network modeling for the test data and total data using different training algorithms.

According to the results of this study, it can be stated that GDBP training algorithm with 8 hidden layers neurons, BFGS with 38 neurons, Rprop with 35 neurons, SCGBP with 33 neurons, and LMBP with 38 neurons have the best performance, respectively.

To find the optimized parameters in RBF neural network, at first, MSE was calculated for obtaining the maximum number of network neurons. It was observed that with the increased maximum number of neurons, MSE firstly reduced, but after 29 neurons, MSE increased. Next, to find a Spread parameter showing the variance, the maximum neuron value of 29 was selected, and MSE was found for various Spreads. Correspondingly, it was observed that with increasing Spread, MSE firstly reduced, but after 64, it increased. MSE and R^2 values for the optimized parameters (Goal=0, MaxNeuron=29, Spread=64) were obtained as 9.98% and 0.84, respectively.

According to the results indicated in tables 1 and 2 and the RBF network, it can be stated that neural network with training algorithms of BFGS with R^2 of 0.91 (R=0.96) and RMSE of 7.40% (MSE=2.72%) has the best performance in the evaluation ESAK. Therefore, in order to estimate ESAK with high accuracy, a two-layer perceptron neural network along with BFGS training algorithms as well as 38 hidden layer neurons can be used. Figure 3 shows the results of the BFGS training algorithm for all the data

[DOI: 10.52547/ijrr.20.3.21]

with a good correspondence; however, it is not complete. However, the changes were insignificant, and it can be ensured that the network has not overfitted, so the correlation coefficient between output and target is 0.96, indicating a high correlation between these two. Furthermore, the RMSE for all the data was obtained as 0.67 mGy, which is equal to 7.40%. Furthermore, it can be seen that, except in some limited instances, the error in other samples is insignificant. The mean and standard deviation in the histogram diagram were also obtained as -0.05 mGy and 0.67 mGy, respectively, and only 3 outlier data exist with an error of more than 2 mGy. In addition, the highest error distribution was around zero. Figure 4 shows the results of the BFGS training algorithm for the test data. According to Figure 4, RMSE for the test data was obtained as 0.81 mGy, which is equal to 8.93%, and the correlation coefficient between output and target was also 0.95, showing a high correlation. Moreover, the mean and standard deviation in the histogram were -0.18 mGy and 0.80 mGy, respectively.

Figure 5 shows the performance diagram of the BFGS training algorithm, the performance index of which was MSE, and the early stopping method was applied in the training process in order to prevent overfitting of the network. The diagram represents the best performance regarding the validation of the data on epoch 35 with a value obtained as 0.56. The network started data overfitting from epoch 35, and despite the decreased MSE values of training data, the MSE value of the validation data increased in the next epochs. Therefore, the network was stopped at this epoch and its weights were used as the final weights.

Table 1. RMSE (root mean square error) and R^2 (coefficient of determination) of GDBP and BFGS algorithms. (mGy: milligray).

GDBP algorithm					BFGS algorithm				
Neuron no.	R^2 All Data	R^2 Test Data	RMSE All Data (mGy)	RMSE Test Data (mGy)	Neuron no.	R^2 All Data	R^2 Test Data	RMSE All Data (mGy)	RMSE Test Data (mGy)
5	0.72	0.72	1.19	1.45	**32**	0.91	0.82	0.69	0.96
6	0.69	0.65	1.25	1.29	**33**	0.88	0.83	0.80	0.83
7	0.74	0.67	1.16	1.34	**34**	0.89	0.85	0.74	0.84
8	0.80	0.86	1.04	1.05	**35**	0.88	0.87	0.77	0.99
9	0.70	0.73	1.25	0.95	**36**	0.83	0.78	0.93	1.10
10	0.76	0.78	1.12	0.98	**37**	0.86	0.82	0.85	0.95
11	0.76	0.51	1.14	1.37	**38**	0.91	0.90	0.67	0.81

Table 2. RMSE (root mean square error) and R^2 (coefficient of determination) of SCGBP, Rprop, and LMBP. algorithms. (mGy: milligray)

	SCGBP algorithm				Rprop algorithm				LMBP algorithm			
Neuron no.	R^2 All Data	R^2 Test Data	RMSE All Data (mGy)	RMSE Test Data (mGy)	R^2 All Data	R^2 Test Data	RMSE All Data (mGy)	RMSE Test Data (mGy)	R^2 All Data	R^2 Test Data	RMSE All Data (mGy)	RMSE Test Data (mGy)
32	0.87	0.89	0.95	0.74	0.82	0.80	0.96	0.84	0.88	0.82	0.80	1.21
33	0.90	0.86	0.72	0.82	0.87	0.77	0.82	1.17	0.85	0.79	0.91	1.29
34	0.86	0.82	0.87	1.07	0.86	0.85	0.86	0.89	0.88	0.82	0.78	1.10
35	0.88	0.86	0.78	0.90	0.88	0.82	0.78	1.06	0.86	0.78	0.87	1.20
36	0.84	0.84	0.90	0.91	0.88	0.80	0.64	1.19	0.88	0.78	0.80	1.21
37	0.85	0.82	0.87	0.94	0.86	0.77	0.85	1.16	0.86	0.74	0.86	1.17
38	0.89	0.85	0.75	0.91	0.82	0.80	0.95	1.12	0.90	0.91	0.70	0.77

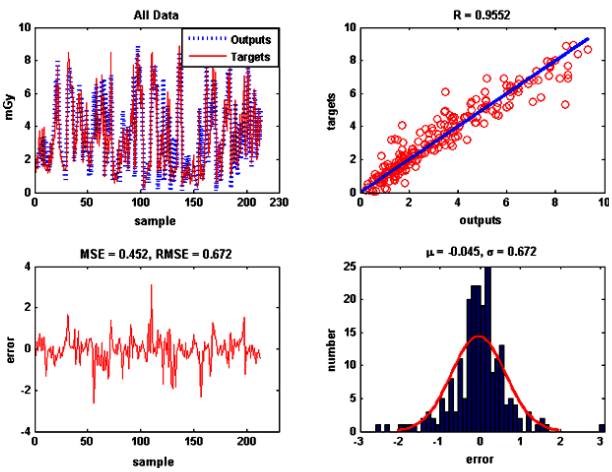

Figure 3. The results of two-layer perceptron neural network with the BFGS training algorithm and 38 hidden layer neurons for all the data.

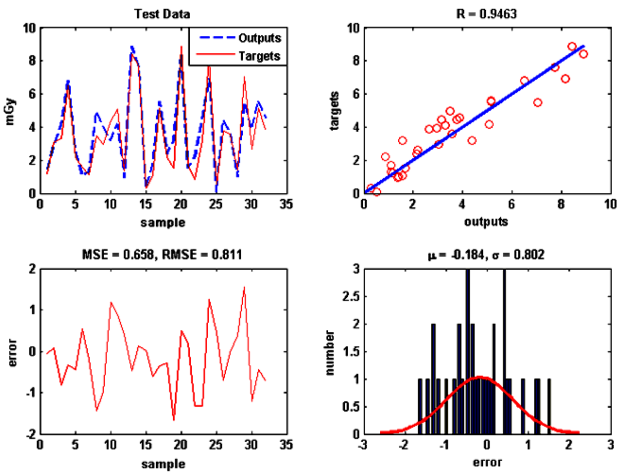

Figure 4. The results of two-layer perceptron neural network with the BFGS training algorithm and 38 hidden layer neurons for the test data.

[DOI: 10.52547/ijrr.20.3.21]

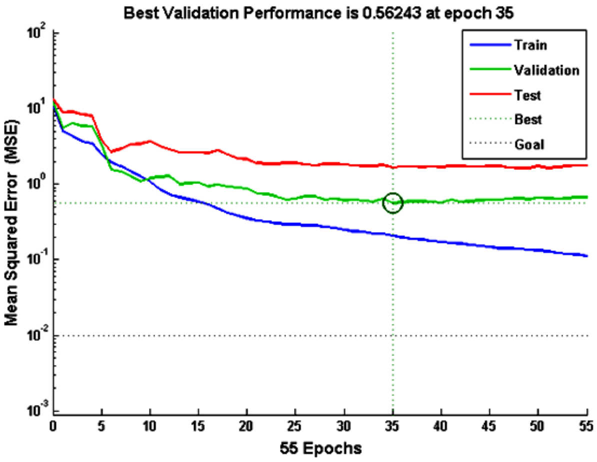

Figure 5. Performance diagram of the most optimal neural network (BFGS) in ESAK modeling.

DISCUSSION

In this study, the ability of ANN to predict the ESAK in mammography screening was investigated to determine the absorbed dose of the breast in the absence of the measurement tools. According to the obtained results, ANN with BFGS training algorithm showed a good performance in predicting the ESAK values. Thus, using this method, the limitations in the process of preparing dosimeters are removed and with the information of individuals as well as the parameters related to the mammography unit, it would be possible to estimate MGD accurately before exposing the patient to radiation.

By comparing the results of this research, it was shown that the modified back-propagation algorithms yield better results compared to GDBP and RBF neural networks, which is in agreement with the results of the Bayram *et al.* study [22]. Notably, the BFGS algorithm is performed based on Newton's method, but it is not needed to compute second-order partial derivatives, and the Hessian matrix is updated approximately in each iteration of the algorithm. The BFGS algorithm has also been recognized as the most successful algorithm among published studies [23] that used quasi-newton methods and is suitable for training small networks.

Hagan *et al.* in 2014 have revealed that if the number of hidden layer neurons is low, the network cannot detect the input-output relationship, and if it is too high, the network begins to memorize the pattern in order to perform well during training; however, it has a poor performance for test data and lacks generalizability in this regard [17]. At this work, the RMSE and the correlation coefficient of the test data for the BFGS algorithm were obtained as 8.93% and 0.95, respectively, which are very close to the results of all data. Correspondingly, it can be inferred that the number of hidden neurons is correctly selected and the network is not overfitted and extrapolated.

In another study, Mohammadi *et al.* [24] have evaluated the MGD during mammography using the two methods of TLD measurement and Monte Carlo simulation; their results showed that the absorbed dose difference rate in breast tissue varied from 5.70% to 17%. However, the present study predicted a MGD with a dose difference rate of 2.72% compared to the measured values.

Most previous studies [4,9,16] have focused on a specific breast model, including skin thickness and breast radius, while ANN allows for more complexity and parameters. This can be considered as an advantage over recent studies that were more sophisticated and provided more realistic breast models for dosimetry calculations. In a study conducted on the estimation of the MGD using MLP neural network using different training algorithms such as LMBP, BFGS, and SCGBP, Čeke *et al.* [25] have selected LMBP training algorithm with a correlation coefficient of 0.85 as the optimal neural network. However, in the present study, ANN showed a correlation coefficient of 0.96. Massera *et al.* [15] in their study have used ANN with Keras and Scikit-learn libraries for regression of dosimetry values applied to mammography, which showed a good performance with a 3% error for predicting the calculated values of both MGD and ESAK. Though, in the present study, the error rate was calculated as 2.72%.

CONCLUSION

In the present research, ANN has successfully found the complex pattern involved between ESAK and various parameters, so it can be considered as an effective method among different protocols and as an alternative or complementary method to other ESAK estimation techniques such as parametric equations and polynomial fit. In addition, new parameters that may play an important role in the value of ESAK in future studies can be added to the neural network along with more data in this regard as well as other preprocessing methods that can be used for the network development.

ACKNOWLEDGEMENTS

The authors would like to send their appreciations to the department of medical physics, School of medicine at Shahid Beheshti University of Medical Sciences (Tehran, Iran). This article has been extracted from the thesis written by Mr. Nabipour in School of Medicine Shahid Beheshti University of Medical Sciences. (Registration No:326).

Ethical approve: This study does not contain any study with human participants or animals performed

[DOI: 10.52547/ijrr.20.3.21]

by any of the authors.

***Conflicts of interest*: Declared none.**

***Funding*: This work was supported by the school of medicine at Shahid Beheshti University of Medical Sciences (Tehran, Iran). grant [NO-326].**

***Authors' contribution*:** Principal Investigator and study design: M.R.D; Methodology: M.N, M.R.D, A.A.A and N.S ; Data analysis: M.N, M.R.D, A.A.A; Interpretation of the findings: M.N, M.R.D, A.A.A; Writing and preparing manuscript: M.N, M.R.D; Reviewing and approving the final version for publication: All authors.

REFERENCES

1. Howlader N, Noone AM, Krapcho M (2021) SEER Cancer Statistics Review. National Cancer Institute, posted to the SEER website. Cited 2021 Apr 19. Available from: https://seer.cancer.gov/csr.
2. gco.iarc.fr [internet]. International Agency for Research on Cancer. Posted to The Global Cancer Observatory Website. 2020 [cited 2021 March 11]. Available from: https://gco.iarc.fr/today/home
3. Iranmakani S, Mortezazadeh T, Sajadian F, Ghaziani MF, Ghafari A, Khezerloo D, *et al.* (2020) A review of various modalities in breast imaging: technical aspects and clinical outcomes. *Egypt J Radiol Nucl Med, **51**: 57*.
4. Dance DR, Skinner CL, Carlsson GA (1999) Breast dosimetry. *Appl Radiat Isot, **50**(1): 185–203*.
5. Osteras BH, Skaane P, Gullien R, Martinsen ACT (2018) Average glandular dose in paired digital mammography and digital breast tomosynthesis acquisitions in a population based screening program: Effects of measuring breast density, air kerma and beam quality. *Phys Med Biol, **63**(3): 035006*.
6. Yaffe MJ and Mainprize JG (2010) Risk of radiation-induced breast cancer from mammographic screening. *Radiology, **258**(1): 98–105*.
7. Nosratieh A, Hernandez A, Shen SZ, Yaffe MJ, Seibert JA, Boone JM (2015) Mean glandular dose coefficients (DgN) for x-ray spectra used in contemporary breast imaging systems. *Phys Med Biol, **60** (18): 7179–90*.
8. Delis H, Spyrou G, Panayiotakis G, Tzanakos G (2005) a Monte Carlo simulation program for dose related studies in mammography. *Eur J Radiol, **54(3):** 371–6*.
9. Nigapruke K, Puwanich P, Phaisangittisakul N, Youngdee W (2010) Monte Carlo simulation of average glandular dose and an investigation of influencing factors. *J Radiat Res, **51**(4): 441–8*.
10. Kave Z, Deevband MR, Ghorbanee M (2018) Effect of heterogeneous breast tissue on average glandular dose in mammography using simulation. *Q J Sch Med Shahid Beheshti Univ Med Sci, **42** (1): 34–9*.
11. Suleiman ME, Brennan PC, McEntee MF (2017) Mean glandular dose in digital mammography: a dose calculation method comparison. *J Med Imaging (Bellingham), **4**(1): 013502*.
12. Boone JM (1999) Glandular breast dose for monoenergetic and high-energy X-ray beams. *Radiology, **213**: 23–37*.
13. Stanton L, Villafana T, Day JL, Lightfoot DA (1984) Dosage evaluation in mammography. *Radiology, **150**(2): 577–84*.
14. Alvar AA, Deevband MR, Ashtiyani M (2017) Neutron spectrum unfolding using radial basis function neural networks. *Appl Radiat Isot, **129**: 35–41*.
15. Massera RT and Tomal A (2021) Breast glandularity and mean glandular dose assessment using a deep learning framework: Virtual patients study. *Phys Medica, **83**: 264–77*.
16. Dance DR, Skinner CL, Young KC, Beckett JR, Kotre CJ (2000) Additional factors for the estimation of mean glandular breast dose using the UK mammography dosimetry protocol. *Phys Med Biol, **45**(11): 3225–40*.
17. Hagan MT, Demuth HB, Beale MH, De Jesús O (2014) Neural Network Design; 2nd Edition, eBook, Frisco, Texas.
18. Yee PV and Haykin SS (2001) Regularized radial basis function networks: theory and applications. John Wiley.
19. Haykin S (1999) Neural networks: a comprehensive foundation second edition. Pearson Education.
20. Hartman EJ, Keeler JD, Kowalski JM (1990) Layered neural networks with Gaussian hidden units as universal approximations. *Neural Comput MIT Press, **2**(2): 210–5.*
21. Buhmann MD (2003) Radial basis functions theory and implementations. Cambridge University Press.
22. Bayram S, Ocal Me, Laptali Oral E, Atis Cd (2016) Comparison of multi-layer perceptron (MLP) and radial basis function (RBF) for construction cost estimation: The case of Turkey. *J Civ Eng Manag. Taylor & Francis, **22**(4): 480-490*.
23. Avriel M (2003) Nonlinear programming: analysis and methods. Dover Publications.
24. Mohammadi A, Faghihi R, Mehdizadeh S, Hadad K (2005) Total absorbed dose of critical organs in mammography, assessment and comparison of Monte-Carlo method and TLD. *Biomed Tech, **50**(1): 393–4*.
25. Čceke D, Kunosic S, Kopric M, Lincender L (2009) Using neural network algorithms in prediction of mean glandular dose based on the measurable parameters in mammography. *Acta Inform Medic, **17**(4): 194–7*.

[DOI: 10.52547/ijrr.20.3.21]