



Enhancement Of Breast Cancer Diagnosis Accuracy With Deep Learning

İlker Yıldız^{1*}, Alper Talha Karadeniz²

¹Bolu Abant İzzet Baysal University, Bolu Technical Science Vocational School, Computer Technologies Department, Bolu, Turkey (ORCID: 0000-0002-1575-2673)

²Bolu Abant İzzet Baysal University, Bolu Technical Science Vocational School, Computer Technologies Department, Bolu, Turkey (ORCID: 0000-0003-4165-3932)

(This publication has been presented orally at HORA congress.)

(First received 1 August 2019 and in final form 25 October 2019)

(DOI: 10.31590/ejosat.638428)

REFERENCE: Yıldız, İ. & Karadeniz, A.T. (2019). Enhancement Of Breast Cancer Diagnosis Accuracy With Deep Learning. *European Journal of Science and Technology*, (Special Issue), 452-462.

Abstract

Breast cancer is a highly fatal disease that is very prevalent among the female population. In this study, a new type of approach is proposed with the aim of improving the accuracy of breast cancer diagnosis, an important problem of our present time, by means of deep learning, one of the techniques in machine learning. In the designed method, the original data set of Breast Cancer Wisconsin being available in the Irvine Machine Learning Repository of University of California was used. Within this data set, there were 699 data consisting of 10 independent variables and 1 dependent variable. The complete utilization of the entire data set was ensured by correction of 16 incorrect data. A normalization process was applied in the data set for the purpose of reducing the time required for learning process. The used data set was allocated as 80% for training, 10% for validation, and 10% for testing. An artificial neural network was designed for the deep learning model. The neural network was set up of a total of 5 layers which were an input layer with 10 neurons, 3 hidden layers with 1000 neurons for each layer, and an output layer with 3 neurons. The software, developed for implementation was written by using Spyder which is an interactive development environment for Python programming language. In addition, Keras neural network API was used. The performance of the model was evaluated with Confusion Matrix and ROC (Receiver Operating Characteristic) analysis. According to the test data obtained at the end of the training, it was observed that the implemented model provided successful results. It is considered that the proposed method will contribute to the improvement of breast cancer diagnosis accuracy.

Keywords: Deep Learning, Breast Cancer, Wisconsin Data Set, Classification.

Derin Öğrenme İle Meme Kanseri Tanısının Doğruluğunun Geliştirilmesi

Öz

Meme kanseri, kadınlarda çok sık görülen ve ölümlere neden olan hastalıklardan biridir. Bu çalışmada makina öğrenmesi tekniklerinden biri olan derin öğrenme metodu ile günümüzün büyük bir problemi olan meme kanseri tanısının doğruluğunun geliştirilmesi amacıyla yeni bir yaklaşım ortaya konulmuştur. Tasarlanan yöntemde, literatürde yer alan University of California, Irvine Makina öğrenmesi veri havuzunda bulunan Breast Cancer Wisconsin orjinal veri seti kullanılmıştır. Bu veri setinde 10 adet bağımsız değişken, 1 adet bağımlı değişkenden oluşan 699 veri mevcuttur. Bu verilerden bozuk olan 16 adet veri düzeltilerek tüm veri

* Corresponding Author: Bolu Abant İzzet Baysal University, Bolu Technical Science Vocational School, Computer Technologies Department, Bolu, Turkey (ORCID: 0000-0002-1575-2673) ilkeryildiz@ibu.edu.tr

setinin kullanılması sağlanmıştır. Veri setinin öğrenme süresinin azaltılması amacıyla normalizasyon işlemi yapılmıştır. Kullanılan veri seti, % 80 eğitim için, %10 değerlendirme ve %10 test için ayrılmıştır. Derin öğrenme modeli için bir yapay sinir ağı tasarlanmıştır. Sinir ağı 10 nöronlu giriş katmanı, 1000'er nöronlu 3 adet gizli katman ve 3 nöronlu çıkış katmanı olmak üzere toplam 5 katmandan oluşturulmuştur. Uygulamada geliştirilen yazılım Python programlama dili için, etkileşimli geliştirme ortamı olan Spyder ile kodlanmıştır. Keras sinir ağı API' si kullanılmıştır. Oluşturulan modelin performansı Confusion Matrix ve ROC (Receiver Operating Characteristic) analizi ile değerlendirilmiştir. Eğitim sonunda elde edilen test verilerine göre gerçekleştirilen modelin başarılı sonuçlar verdiği görülmüştür. Önerilen yöntemin meme kanseri tanısının doğruluğunun geliştirilmesine katkıda bulunacağı değerlendirilmektedir.

Anahtar Kelimeler: Derin Öğrenme, Meme Kanseri, Wisconsin Veri seti, Sınıflandırma.

1. Introduction

Being a disease with fatal outcomes, breast cancer is considered as the second most common cancer type among females around the world. According to the data provided by World Cancer Research Fund International, 2 million new cases were observed in 2018, and this number is gradually increasing [1]. In Turkey, twenty-five thousand patients are diagnosed with breast cancer each year. This rate continues to increase in developing countries. For this reason, the importance of accurate and early diagnosis has become higher than ever. Moreover, the rate of machine learning usage in diagnosis of diseases within the field of medicine is growing. These techniques provide assistance to physicians in achieving an early and accurate diagnosis. In this study, a new type of approach is proposed with the purpose of improving the accuracy of breast cancer diagnosis by means of one of the techniques found in the concept of machine learning that is called deep learning.

Artificial intelligence enables the machines to operate by imitating the human capabilities of learning and decision making. Machine Learning (ML) is a discipline of artificial intelligence that ensures the software to estimate results with better accuracy, without the need to write explicit codes to perform the task mentioned. In order to design a model using the conventional machine learning techniques, the feature vector must be extracted initially. Moreover, there is a requirement of expert supervision in these techniques, and the crude data cannot be directly used. However with deep learning, the supervision of an expert is not needed, and the direct use of crude data is possible [2, 3]. Deep learning is one of the many methods found within machine learning. Deep learning methods are utilized in machine learning in terms of quick learning and implementation of large and complex data. Deep learning is widely used in many software disciplines such as computer vision, speech and sound processing, natural language processing, robotics, bioinformatics, computer games, search engines, energy generation, automotive industry, aviation, manufacturing, online advertising and financing, and etc. [4]. It is known that deep learning provides highly successful results in processes of estimation and classification.

Today, there are quite many researches present in the literature regarding the subject of breast cancer diagnosis. In addition, there were also researches that were conducted using the data set of Wisconsin Breast Cancer. Banerjee et al. presented a comparative study of different Ensemble Learning techniques using the Wisconsin Breast Cancer dataset [5]. Ghosh et al. performed different statistical and deep learning analyses of the Wisconsin Breast Cancer Database to improve accuracy in the detection and classification of breast cancer [6]. Mumtaz et al. performed Evaluation of Three Neural Network Models using Wisconsin Breast Cancer Database [7]. Ashraf proposed an Iterative Weighted k-NN for Constructing Missing Feature Values in Wisconsin Breast Cancer Dataset [8]. In [9], the feature selection and feature extraction techniques were combined together by using deep learning in order to predict the outcome of the breast cancer.

In this study, a newer method was proposed to enhance the accuracy of the diagnosis of breast cancer by means of deep learning. In the designed method, the original data set of Breast Cancer Wisconsin, a data set used in diagnosis of breast cancer that took place in the literature, was used. By carrying out a classification process, it was ensured that the method gave results as benign or malignant at output according to the values of patient groups in the data set. Confusion Matrix metrics and ROC (Receiver Operating Characteristic) analysis were utilized in the evaluation of model performance.

In the second part of the paper, the experimental studies and the methods used are given. Section 3 involves the test results obtained from the experimental studies and the evaluation of the study is given in section 4.

2. Material and Method

In this study, the processes conducted within the method that was designed with deep learning with the aim of enhancing the accuracy of the diagnosis of breast cancer were explained as follows.

The Flow Chart of the designed model can be seen in Figure 1.

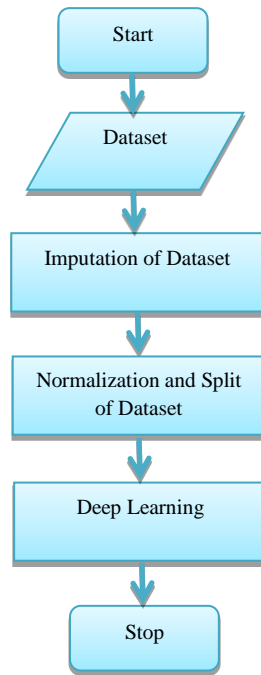


Figure 1. Flow Chart of the model

2.1. Breast Cancer Wisconsin Original Data Set

In the method proposed for breast cancer diagnosis, the original data set of Breast Cancer Wisconsin available in the Machine Learning Repository of University of California, Irvine, and which also took its place in the literature, was used [10]. The data set only contains numerical values. Within this data set, a total of 699 data were available, consisting of 458 benign and 241 malignant patient groups. The data set was constituted of 10 independent variables and 1 dependent variable. 16 of these data had missing feature values. The feature names and values of the data found in the data set is provided on Table 1. A snapshot from the data set can be seen in Figure 2.

Table 1. Breast Cancer Wisconsin original data set feature names and values that are used in this study

Attribute	Domain
Sample code number	Id No
Clump Thickness	1 - 10
Uniformity of Cell Size	1 - 10
Uniformity of Cell Shape	1 - 10
Marginal Adhesion	1 - 10
Single Epithelial Cell Size	1 - 10
Bare Nuclei	1 - 10
Bland Chromatin	1 - 10
Normal Nucleoli	1 - 10
Mitoses	1 - 10
Class	2 for benign, 4 for malignant

1000025	5	1	1	1	2	1	3	1	1	2
1002945	5	4	4	5	7	10	3	2	1	2
1015425	3	1	1	1	2	2	3	1	1	2
1016277	6	8	8	1	3	4	3	7	1	2
1017023	4	1	1	3	2	1	3	1	1	2
1017122	8	10	10	8	7	10	9	7	1	4
1018099	1	1	1	1	2	10	3	1	1	2
1018561	2	1	2	1	2	1	3	1	1	2
1033078	2	1	1	1	2	1	1	1	5	2
1033078	4	2	1	1	2	1	2	1	1	2
1035283	1	1	1	1	1	1	3	1	1	2
1036172	2	1	1	1	2	1	2	1	1	2
1041801	5	3	3	3	2	3	4	4	1	4
1043999	1	1	1	1	2	3	3	1	1	2
1044572	8	7	5	10	7	9	5	5	4	4
1047630	7	4	6	4	6	1	4	3	1	4
1048672	4	1	1	1	2	1	2	1	1	2
1049815	4	1	1	1	2	1	3	1	1	2
1050670	10	7	7	6	4	10	4	1	2	4
1050718	6	1	1	1	2	1	3	1	1	2
1054590	7	3	2	10	5	10	5	4	4	4
1054593	10	5	5	3	6	7	7	10	1	4
1056784	3	1	1	1	2	1	2	1	1	2
1057013	8	4	5	1	2	?	7	3	1	4
1059552	1	1	1	1	2	1	3	1	1	2
1065726	5	2	3	4	2	7	3	6	1	4
1066373	3	2	1	1	1	1	2	1	1	2

Figure 2. A snapshot of dataset

2.2. Imputation of Dataset

The training of a model with a data set containing missing values may substantially affect the quality of deep learning model. For this reason, the utilization of entire data set in training was ensured by correction of 16 incorrect data found in the data set with a statistical missing value analysis technique called Mean Imputation. This method functions by calculating the mean value of readily available values in a column, and then substituting the missing values in each column independently from each other [11].

The method does not have an effect upon the class variable of “Sample Code Number” feature located in the data set. In order to reduce the dimensionality of the data set and to prevent the addition of trivial features, this column was excluded from the data set.

2.3. Normalization of Dataset

Application of normalization on the crude data increases the data efficiency. A normalization process between the ranges of 0-1 was applied in the data set for the purpose of reducing the long learning period caused by the size of the data set. The MinMaxScaler method was used in this process as shown in Equation 1 [12].

$$z = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{1}$$

Here, z is the normalized data, x is the input value, min (x) is the smallest number in the input set, max (x) is the largest number in the input set.

2.4. Splitting the Dataset

The data used in deep learning are separated in 3 clusters as validation, training and testing. The allocation of available data among these three data sets is vital for the objectivity of the success. As a result of various tests, the data set in the suggested model was allocated as 80% (559 data) for training, 10% (70 data) for validation, and 10% (70 data) for testing. The process of allocation is shown in Figure 3. Cross validation method was used in the implementation of this process [13].



Figure 3. Splits of dataset.

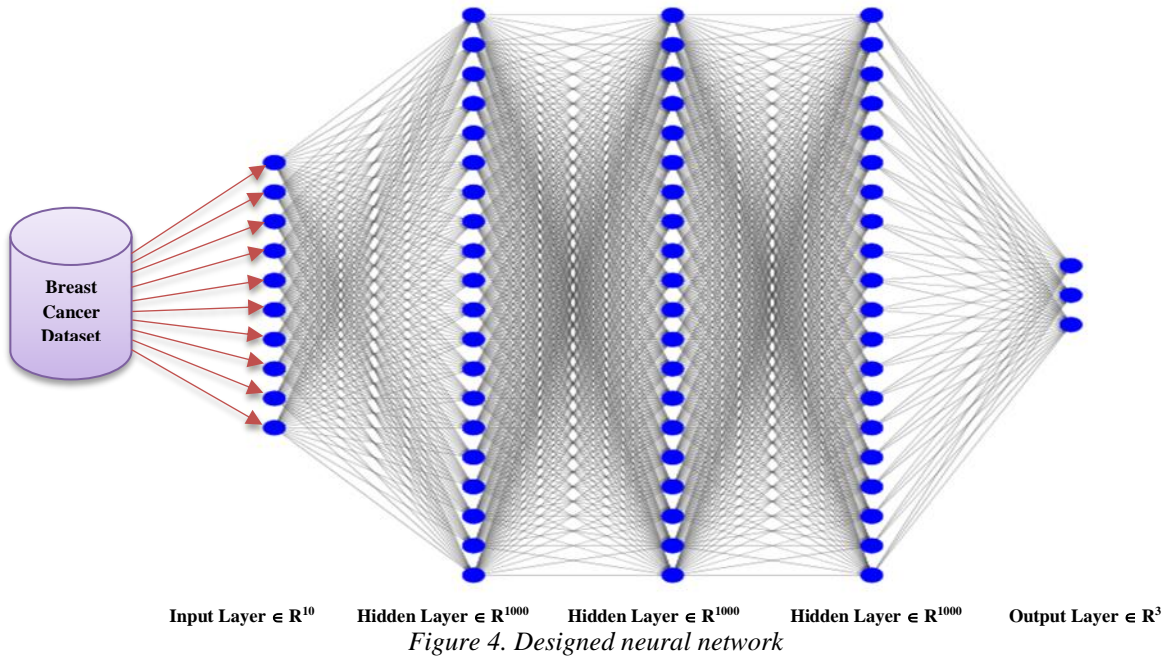
2.5. Software

The software which was developed for the study was written using Spyder for Python programming language, an interactive development environment capable of advanced editing, interactive testing, debugging and introspection. In addition, Keras neural network API was used for deep learning in the developed method. Keras is a high level neural network API, supporting Python. It is

able to convert the results rapidly, highly modular, minimalist, and has extensible features. Keras supports CNN (Convolutional Neural Network), RNN (Recurrent Neural Network) and a combination of the Neural Networks [14].

2.6. Neural Network Model

Keras Sequential model was used for implementation. An artificial neural network with 5 layers was designed by carrying out several tests on determination of the Artificial Neural Network parameters. Within this neural network, there is an input layer with 10 neurons, 3 hidden layers consisting of 1000 neurons for each layer, and a 3-neuron output layer. The designed neural network is shown in Figure 4.



The number of neurons in the input layer of the neural network is 10 in that there are 10 different inputs in the data set used for implementation. The value which the data set may get in the output layer may be 2 or 4 (2 for benign, 4 for malignant). Because that sigmoid activation function is used at the output layer, the number 4 is represented in 3 bits as a binary value. For this reason, the number of neurons at the output layer of the network was designed to be three.

2.6.1. Activation Functions of the Neural Network:

The activation functions used in the layers of created neural network are described as follows.

2.6.1.1. ReLU (Rectified Linear Unit) Activation Function

ReLU (Rectified Linear Unit) Activation Function was used in the input layer and hidden layers of the neural network. ReLU Activation Function can be seen in Figure 5. ReLU is an activation function that recently gained popularity for its practicality in deep learning. It enables the neural network to learn faster. The function zeroes negative values [15]. The mathematical expression of the function is provided in Equation 2.

$$f(x) = \max(0, x) \tag{2}$$

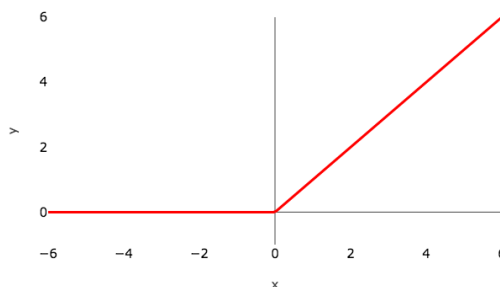


Figure 5. ReLU Activation Function

2.6.1.2. Sigmoid Activation Function

Sigmoid activation function was used at the output layer of the neural network. Sigmoid Activation Function is a function that gets a value between the ranges of (0, 1) as seen in Figure 6. The mathematical expression of the function is provided in Equation 3.

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{3}$$

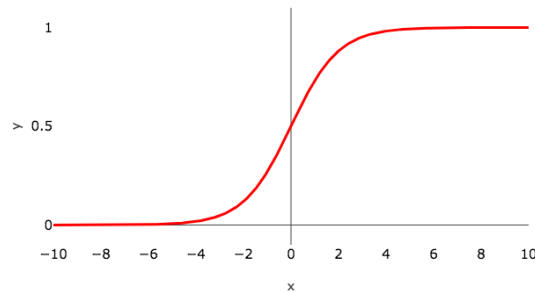
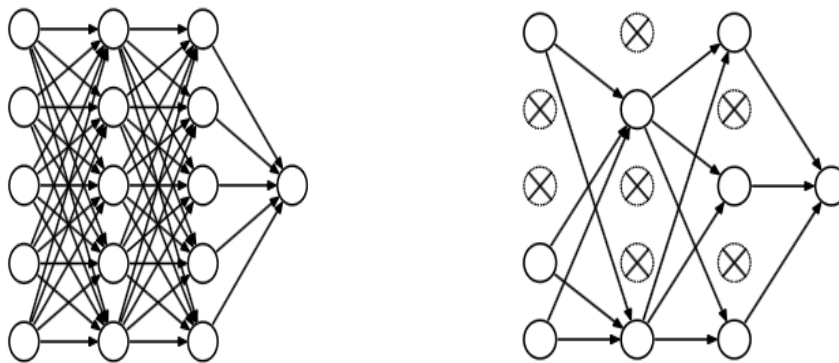


Figure 6. Sigmoid Activation Function

2.7. Dropout

Dropout is one of the methods that is used to prevent memorization. In each iteration, it randomly removes a number of neurons from a layer at a specified rate. The process of dropout is described in Figure 7. Crossed units have been dropped out of the network.



a) A standard neural network with 2 hidden layers.

b) Applying dropout to the network

Figure 7. Dropout Neural Network Model [16].

Figure 7 a shows a standard neural network with 2 hidden layers and figure 6 b shows after applying dropout to the network.

Regularization techniques are also used to increase performance in Dropout method by prevention of overfitting. Regularizers L2 formula is also used along with the Dropout process in the designed model. In this way, overfitting during the training of the network is minimized.

2.8. Optimization

The learning process in deep learning applications is essentially a problem of optimization. Optimization techniques are used to find the optimum value in solving non-linear problems. RMSprop, adagrad, adadelata, adam, adamax and such other optimization algorithms are widely used in deep learning applications. Moreover, there are differences between each of these algorithms in terms of performance and speed. In this study, the optimization algorithm of Adaptive Moment Optimization (Adam) was used.

2.8.1. Adam Optimization

The Adam algorithm is simple to implement, has fast learning time, is computationally efficient, requires little memory, and is well suited for problems with large data or parameters. [17]. The Adam optimization algorithm combines with both momentum and RMSprop algorithms. The equations of the algorithm are given below [18].

For each Parameter ω^j

(*j* subscript dropped for clarity)

$$v_t = \beta_1 * v_{t-1} - (1 - \beta_1) * g_t \quad (4)$$

$$s_t = \beta_2 * s_{t-1} - (1 - \beta_2) * g_t^2 \quad (5)$$

$$\Delta\omega_t = -\eta \frac{v_t}{\sqrt{s_t + \epsilon}} * g_t \quad (6)$$

$$\omega_{t+1} = \omega_t + \Delta\omega_t \quad (7)$$

Here, η : **Initial learning rate**

g_t : **Gradient at time t along ω^j**

v_t : **Exponential Average of gradients along ω^j**

s_t : **Exponential Average of squares of gradients along ω^j**

β_1, β_2 : **Hyperparameters**

Here, the exponential average of the gradient and the squares of the gradient computes for each parameters (Eq 4, and Eq 5). To determine the learning step, learning rate is multiplied by the exponential average of the gradient and it is divided by the root mean square of the exponential average of square of gradients (Eq 6). Then update is done. In practice the hyper parameter values for $\beta_1 = 0.9$ and $\beta_2 = 0.99$. Epsilon is (ϵ) is a very small number (10^{-8}) to avoid dividing by zero.

2.9. Loss Function

Loss function is a type of function that measures both the error rate and performance of a designed model. In deep learning, the last layer of a neural network is the layer where the loss of function is defined. In deep learning applications, the function calculates the dissimilarity between the estimation of the designed model and the required real value. In case that a model with good estimation capability is designed, the difference between the real value and estimated value will be lower. In other words, its loss value will be lower. An output of higher loss value indicates that the designed model contains flaws. However, in a finely designed model, a loss value near zero must be expected. In the literature, there are various loss functions such as mean squared error, mean absolute percentage error, mean squared logarithmic error, hinge, logcosh, sparse categorical cross entropy, binary cross entropy, kullback leibler divergence, poisson, cosine proximity, and many others. In this study, the sparse categorical cross entropy loss function was used.

2.9.1. Sparse Categorical Cross-entropy Loss Function

Computes the categorical cross-entropy between predictions and targets. This loss function is a choice for multi-class classification problems. The mathematical equation is given in Equation 8 [19].

$$L_i = -\sum_j t_{i,j} \log(p_{i,j}) \quad (8)$$

Here p refers the predictions, t refers the targets, i refers the data point and j refers the class. Here, accuracy metrics are used with this loss function.

2.9.2. Early Stopping

In the models where training is done by iteration with data, the period of learning must be terminated at the right time. Otherwise, if training is not stopped, all of the samples in the data set for training will be memorized by the system. This results in a decrease in the capability of estimation of unknown samples. The time of training termination depends on the choice of algorithm. In case of an early termination, performance of the system will decline in that it could not fully analyze the data. The same outcome will also arise in case of over-training. In case of an overfitting possibility for the program, a parameter of early stopping was defined. In the program, this parameter was defined as 10, meaning that if the neural network returns the same result in 10 consecutive times, the training will be stopped regardless of number of iterations.

2.10. Evaluation of Model Performance

In deep learning, evaluation of model performance is important for the determination of the system performance. The performance of the model realized in this study was evaluated with Confusion Matrix and ROC analysis.

2.10.1. Confusion Matrix

In the classification problem in machine learning, a confusion matrix, is a special table layout that allows the evaluation of the performance of an algorithm. Each row or column of the matrix represents the predicted values, while each row or column represents the values in an actual value [20].

Confusion matrix shows the amounts of true and false estimations of the classification model with the present state of the data. A confusion matrix can be seen in Figure 8. The matrix element varies according to the number of predicted values. The actual values are the output values found in the data set. It gets a positive value for a positive case, and a negative value for a negative case. The objective of the model created with deep learning method is to correctly estimate both the positive and the negative values. The following values are present in the confusion matrix.

True positive (TP) = The number of samples correctly marked as positive.

False positive (FP) = The number of samples incorrectly marked as positive.

True negative (TN) = The number of samples correctly marked as negative.

False negative (FN) = The number of samples incorrectly marked as negative.

Confusion Matrix		Actual Values	
		Positive	Negative
Predicted Values	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Figure 8. Confusion matrix

After the determination of confusion matrix, the performance of the algorithm is evaluated by calculating the values of True Positive Rate, True Negative Rate, Positive Predictive Value, and Negative Predictive Value [21]. In addition, values of Accuracy and F₁ Score are also measured. The mathematical equations are given in Equations 9-14.

True Positive Rate: Also known as Sensitivity and Recall. It measures the proportion of actual positives that are correctly identified.

True Negative Rate: Also known as Specificity. It measures the proportion of actual negatives that are correctly identified.

Positive Predictive Value: Also known as Precision. It is the proportion of positive results that is true positive results.

Negative Predictive Value: It is the proportion of negative results that is true negative results.

Accuracy: It is the ratio of accurate predicts to all predictors in the system.

F₁ Score: It is the harmonic mean of the true positive rate (Recall) and Precision.

$$\text{True Positive Rate} = \frac{TP}{(TP+FN)} \tag{9}$$

$$\text{True Negative Rate} = \frac{TN}{(TN+FP)} \tag{10}$$

$$\text{Positive Predictive Value} = \frac{TP}{(TP+FP)} \tag{11}$$

$$\text{Negative Predictive Value} = \frac{TN}{(TN+FN)} \tag{12}$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{13}$$

$$F_1 \text{ Score} = \frac{2*TP}{(2*TP+FP+FN)} \tag{14}$$

2.10.2. Receiver Operator Characteristic Curve (ROC)

The Receiver Operating Characteristic Curve is a very good method to measure the performance of a classification model. Uses the data obtained in the confusion matrix. ROC analysis investigates and employs the relationship between sensitivity and specificity of a binary classifier [21]. The overall accuracy of the test increases as the ROC curve approaches the upper left corner. [22]. The model performance is determined by looking at the area under the ROC curve (AUC). The AUC value is between 0.5 and 1, which is the best when approaching 1.

3. Findings and Discussion

In this section, the results were provided with the utilization of techniques of Confusion Matrix, ACC, ROC curve and AUC which were used in the performance evaluation of the model designed with deep learning for enhancement of breast cancer diagnosis accuracy.

Epoch was determined as 100 by the results of conducted tests. The confusion matrix of 70 testing data which was obtained as a result of training was shown in Figure 9.

Confusion Matrix		Actual Values	
		Positive	Negative
Predicted Values	Positive	True Positive (TP) 41	False Positive (FP) 1
	Negative	False Negative (FN) 1	True Negative (TN) 27

Figure 9. Confusion Matrix of the designed method.

In the Confusion Matrix, 41 actual malignant data were identified as malignant in the TP value. 1 actual benign datum was determined as malignant in the FP value. 1 actual malignant datum was identified as benign in the FN value. Lastly, 27 actual benign data was identified as benign in the TN value.

The values of Accuracy, AUC, F1 Score, True Positive Rate, True Negative Rate, Positive Predictive Value, and Negative Predictive Value obtained following the training of the deep learning model were presented on Table 2.

Table 2. Calculated Parameters

Parameter	Value
Accuracy	0.9857
F1 Score	0.97619
True Positive Rate	0.9761
True Negative Rate	0.9642
Positive Predictive Value	0.9761
Negative Predictive Value	0.6428
AUC	0.9983

The Receiver Operating Characteristic Curve (ROC), obtained using the confusion matrix and parameter values on Table 2, is shown in Figure 10.

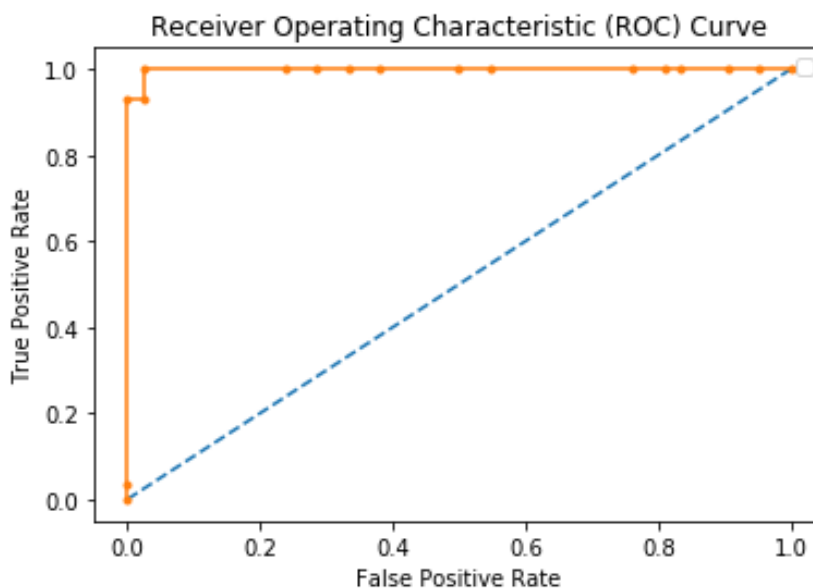


Figure10. ROC Curve of the designed method.

The AUC value was calculated as **0.9983**.

For the purpose of measuring the performance of the proposed model during the phases of training and validation, the following graph in Figure 11 was obtained by measuring the accuracy values (ACC) in each iteration.

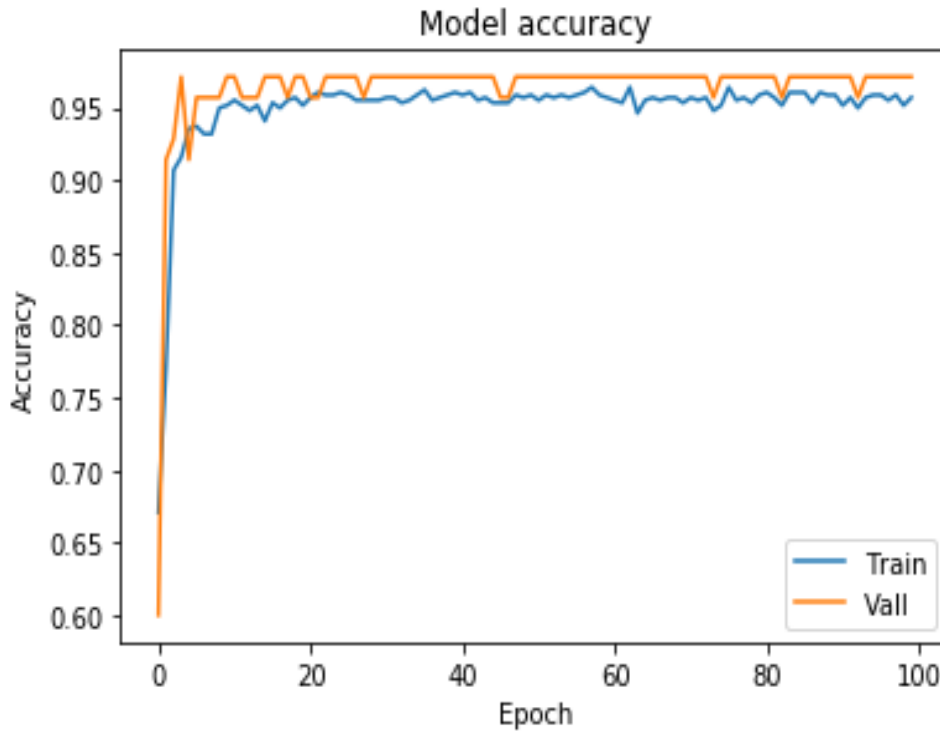


Figure11. Model Accuracy.

The graph, created with the measurement of loss function outputs for each iteration in both the training and the validation phases, is shown in Figure 12.

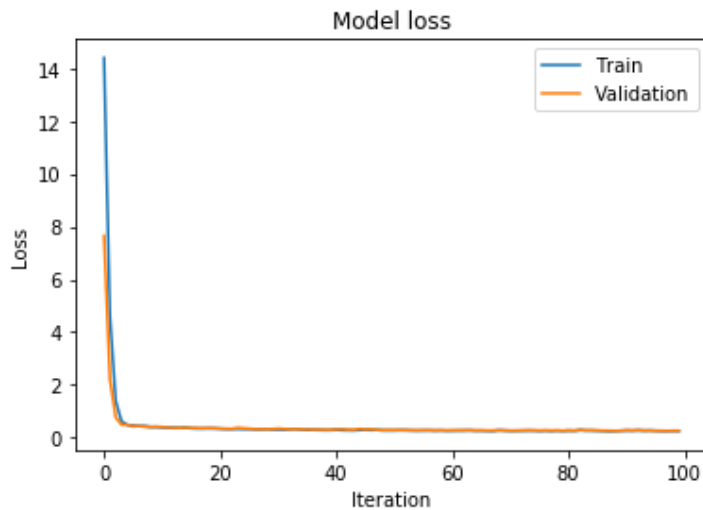


Figure 12. Model Loss

A snapshot of the designed model during the phase of training is shown in Figure 13.

```

559/559 [=====] - 1s 2ms/step - loss: 0.2692 - acc: 0.9696 - val_loss: 0.2614 - val_acc: 0.9429
Epoch 37/100
559/559 [=====] - 1s 2ms/step - loss: 0.2624 - acc: 0.9732 - val_loss: 0.2569 - val_acc: 0.9429
Epoch 38/100
559/559 [=====] - 1s 2ms/step - loss: 0.2654 - acc: 0.9660 - val_loss: 0.2700 - val_acc: 0.9571
Epoch 39/100
559/559 [=====] - 1s 2ms/step - loss: 0.2582 - acc: 0.9714 - val_loss: 0.2596 - val_acc: 0.9429
Epoch 40/100
559/559 [=====] - 1s 2ms/step - loss: 0.2582 - acc: 0.9606 - val_loss: 0.2515 - val_acc: 0.9429
Epoch 41/100
559/559 [=====] - 1s 2ms/step - loss: 0.2517 - acc: 0.9714 - val_loss: 0.2587 - val_acc: 0.9571
Epoch 42/100
559/559 [=====] - 1s 2ms/step - loss: 0.2617 - acc: 0.9696 - val_loss: 0.2626 - val_acc: 0.9571
Epoch 43/100
559/559 [=====] - 1s 2ms/step - loss: 0.2588 - acc: 0.9624 - val_loss: 0.2712 - val_acc: 0.9714
Epoch 44/100
559/559 [=====] - 1s 2ms/step - loss: 0.2526 - acc: 0.9696 - val_loss: 0.2779 - val_acc: 0.9714
Epoch 45/100
559/559 [=====] - 1s 2ms/step - loss: 0.2554 - acc: 0.9678 - val_loss: 0.2478 - val_acc: 0.9429
Epoch 46/100
559/559 [=====] - 1s 2ms/step - loss: 0.2528 - acc: 0.9642 - val_loss: 0.2456 - val_acc: 0.9429

```

Figure 13. A snapshot of the model training.

4. Conclusion

In this study, a new model was suggested with the aim of enhancing the accuracy of the diagnosis of breast cancer with deep learning method. The assessment outcomes of the model that was designed using Breast Cancer Wisconsin Original Data Set were presented in the 3rd Section of this study. The AUC value of the model was measured as **0.9983** and its value of accuracy was measured as **0.9857**. The evaluation results indicated that the proposed model had a high rate of performance. It is thought that this newly developed method will contribute to the improvement of breast cancer diagnosis accuracy, which is a crucial problem of our day.

References

- [1] <https://www.wcrf.org/dietandcancer/cancer-trends/breast-cancer-statistics> (accessed 20.05.2019)
- [2] Montúfar, G.F. (2014). Universal Approximation Depth and Errors of Narrow Belief Networks with Discrete Units, *Neural Computation*, Vol. 26, issue 7, pp 1386-1407.
- [3] Schmidhuber, J. (2015). Deep Learning in Neural Networks: An Overview. *Neural Networks*, Vol. 61, pp.85-117.
- [4] Goodfellow I, Bengio Y, Courville A. (2016). *Deep Learning*, MIT Press.
- [5] Banerjee, C., Paul, S., Ghoshal, M. (2017). A Comparative Study of Different Ensemble Learning Techniques using Wisconsin Breast Cancer dataset, *International Conference on Computer, Electrical & Communication Engineering (ICCECE)*
- [6] Ghosh, S., Hossain, J., Fattah S.A., Shahnaz, C. and Khan, A. I. (2017). Efficient approaches for accuracy improvement of breast cancer classification using Wisconsin database, *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*, Dhaka, pp. 792-797.
- [7] Mumtaz, K., Sheriff, S., Duraiswamy, K.(2009).Evaluation of Three Neural Net-work Models using Wisconsin Breast Cancer Database, *International Conference on Control, Automation, Communication and Energy Conservation*.
- [8] Ashraf, M., Le, K., Huang, X. (2011). Iterative weighted k-nn for constructing missing feature values in Wisconsin breast cancer dataset. In: 2011 3rd International Conference on Data Mining and Intelligent Information Technology Applications (ICMIA), pp. 23–27
- [9] Zhang, D., Zou, L., Zhou, X. and He, F.(2018).Integrating feature selection and feature extraction methods with deep learning to predict clinical outcome of breast cancer, *IEEE Access*, vol. 6, pp. 28936–28944.
- [10] Wisconsin Breast Cancer original dataset [https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+\(original\)](https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(original)) (accessed 10.05.2019)
- [11] Mean Imputation technique. https://chrisalbon.com/machine_learning/preprocessing_structured_data/impute_missing_values_with_means/ (accessed 20.05.2019)
- [12] <https://scikit-learn.org/stable/modules/preprocessing.html> (accessed 10.05.2019)
- [13] https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.cross_validate.html ((accessed 10.05.2019)
- [14] Zhang, T., Zheng, H., Zhang L. (2018). Verification CAPTCHA Based on Deep Learning *Proceedings of the 37th Chinese Control Conference*, Wuhan, China
- [15] Agarap, A.F. (2019).Deep learning using rectified linear units (relu), [Online]. ArXiv: 1803.08375v2 [cs.NE] 7 Feb 2019
- [16] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting, *Journal of Machine Learning Research*, 15(1), pp: 1929–1958.
- [17] Kingma, D. P. and Ba, J. (2015).Adam: A Method for Stochastic Optimization. ArXiv:1412.6980v8 [cs.LG] 23 Jul 2015
- [18] <https://www.jqr.com/article/000505> (accessed 5.05.2019)
- [19] https://lasagne.readthedocs.io/en/latest/modules/objectives.html#lasagne.objectives.categorical_crossentropy (accessed 15.05.2019)
- [20] Powers, D. M. W. (2011). Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation, *Journal of Machine Learning Technologies*, 2 (1), pp.37–63.
- [21] Ting, K. M. (2011). *Encyclopedia of machine learning*. Springer.
- [22] Zweig, M. H. and Campbell, G. (1993) “Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine,” *Clin. Chem.*, vol. 39, no. 4, pp. 561–577.