

South Dakota State University  
**Open PRAIRIE: Open Public Research Access Institutional  
Repository and Information Exchange**

---

Theses and Dissertations

---

2017

# Optimized Multilayer Perceptron with Dynamic Learning Rate to Classify Breast Microwave Tomography Image

Chulwoo Pack  
*South Dakota State University*

Follow this and additional works at: <http://openprairie.sdstate.edu/etd>

 Part of the [Computer Engineering Commons](#), [Computer Sciences Commons](#), and the [Health Information Technology Commons](#)

---

## Recommended Citation

Pack, Chulwoo, "Optimized Multilayer Perceptron with Dynamic Learning Rate to Classify Breast Microwave Tomography Image" (2017). *Theses and Dissertations*. 1696.  
<http://openprairie.sdstate.edu/etd/1696>

This Thesis - Open Access is brought to you for free and open access by Open PRAIRIE: Open Public Research Access Institutional Repository and Information Exchange. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Open PRAIRIE: Open Public Research Access Institutional Repository and Information Exchange. For more information, please contact [michael.biondo@sdstate.edu](mailto:michael.biondo@sdstate.edu).

OPTIMIZED MULTILAYER PERCEPTRON WITH DYNAMIC LEARNING RATE  
TO CLASSIFY BREAST MICROWAVE TOMOGRAPHY IMAGE

BY

CHULWOO PACK

A thesis submitted in partial fulfillment of the requirements for the

Master of Science

Major in Computer Science

South Dakota State University

2017

OPTIMIZED MULTILAYER PERCEPTRON WITH DYNAMIC LEARNING RATE  
TO CLASSIFY BREAST MICROWAVE TOMOGRAPHY IMAGE

This thesis is approved as a creditable and independent investigation by a candidate for the Master of Science in Computer Science degree and is acceptable for meeting the thesis requirements for this degree. Acceptance of this does not imply that the conclusions reached by the candidates are necessarily the conclusions of the major department.

Dr. Sung Y. Shin, Ph.D.  
Thesis Advisor

Date

Dr. Steven Hietpas, Ph.D.  
Head, Department of Electrical Engineering  
and Computer Science

Date

Dean, Graduate School

Date

This thesis is dedicated to all whom suffer from breast cancer.

## ACKNOWLEDGEMENTS

I would like to take this opportunity to thank my advisor, Prof. Sung Shin, who initially convinced me to work in the field of Computer Science. I would also like to thank the members of my thesis committee, Prof. George Hamer, Prof. Ali Salehnia, and Prof. Paul Reynolds, not only for their time, but also for their intellectual contributions.

Besides my thesis committees, I would like to thank Dr. Seong H. Son from ETRI who collaborated with our team by giving his knowledge of electrical engineering to help build an expert system for breast cancer screening. Also, I would like to mention all my team members that I worked with in CCT LAB who fill me with fun and feedback.

Finally, I cannot thank enough my parents and sisters for what they have done for me. Without their help, this thesis could not have been published.

## TABLE OF CONTENTS

LIST OF FIGURES .....	vii
LIST OF TABLES .....	viii
ABSTRACT .....	ix
1. INTRODUCTION .....	1
2. LITERATURE REVIEW .....	5
2.1. COMPUTER AIDED DIAGNOSIS SYSTEM .....	5
2.2. MULTI-LAYER PERCEPTRONS .....	7
3. RELATED WORK .....	9
3.1. FEATURE EXTRATION AND SELECTION .....	9
3.2. CLASSIFICATION .....	10
4. PROPOSED MODEL .....	11
4.1. CLASSIFICATION PHASE .....	12
5. RESULT AND ANALYSIS .....	17
5.1. EVALUATION PHASE .....	17
5.2. RESULTS .....	18
6. CONCLUSION .....	21
LITERATURE CITED .....	23

## ABBREVIATIONS

ANN Artificial Neural Network

CAD Computer Aided Diagnosis

DLR Dynamic Learning Rate

MCC Matthews Correlation Coefficient

MLP Multilayer Perceptron

MRI Magnetic Resonance Imaging

MSE Mean Squared Error

MTI Microwave Tomography Imaging

SVM Support Vector Machine

## LIST OF FIGURES

Figure 1. Diagram of proposed MLP model using DLR .....	3
Figure 2. Expert system: Diagram of CAD system for early detection of breast cancer....	6
Figure 3. The influence of learning rate on the weight changes.....	14
Figure 4. Average value of confusion matrix analysis .....	19
Figure 5. MCC value comparision.....	20



## LIST OF TABLES

Table 1. Confusion matrix of MLP using DLR .....	18
Table 2. Confusion matrix of conventional model .....	19

## ABSTRACT

OPTIMIZED MULTILAYER PERCEPTRON WITH DYNAMIC LEARNING RATE  
TO CLASSIFY BREAST MAMMOGRAM TOMOGRAPHY IMAGE

CHULWOO PACK

2017

Most recently developed Computer Aided Diagnosis (CAD) systems and their related research is based on medical images that are usually obtained through conventional imaging techniques such as Magnetic Resonance Imaging (MRI), x-ray mammography, and ultrasound. With the development of a new imaging technology called Microwave Tomography Imaging (MTI), it has become inevitable to develop a CAD system that can show promising performance using new format of data. The platform can have a flexibility on its input by adopting Artificial Neural Network (ANN) as a classifier. Among the various phases of CAD system, we have focused on optimizing the classification phase that directly affects its performance.

In this paper, we present the optimized Multilayer Perceptron (MLP) binary classifier, which can be plugged into the CAD system, that uses Dynamic Learning Rate (DLR) for alleviating local minima problem. The proposed classifier has an optimized size of neural network so that it will not fall into indeterminate equation problem by having reasonable amount of weights between each perceptron. Also, the proposed model will dynamically assign a learning rate onto each training points in the way that model

earmarks a higher learning rate onto each training points belonging into minority class in order to escape from local minima which is a typical jeopardy of MLP.

In experiment, we evaluate performance of our model with following measures; precision, recall, specificity, accuracy, and Matthews Correlation Coefficient (MCC) and compare them to that of work by Samaneh et al. The results show that our model outperforms existing model not only for the performance such as recall, specificity, accuracy, and precision, but also for the quality, and thus it empowers physicians to make better decision on breast cancer screening in early stage, as it also alleviates the cost burden from the patients.

## 1. INTRODUCTION

Breast cancer is currently the most dominant cancer for women in both the developed and the developing countries. Improving the outcome and survival from breast cancer still remains as the cornerstone of breast cancer control in the modern era [1]. Detecting a tumor at an early stage plays an important role for enhancing survival rate (or reducing mortality), and the CAD system has been helping physicians for detection of breast cancer on mammograms in the United States [2]. The concept of the CAD system is that digitalized medical image processing work by computer could be utilized by physicians or radiologists, but not replace them [2]. A CAD system, as a second opinion, has been investigated in abounding research successfully revealing breast masses and micro-calcification and developed to support the practitioner [6;13]. This should not be misunderstood as the same concept as automated computer diagnosis. Automated computer diagnosis considers a computer as the subject of final decision-making, while medical practitioner, in CAD, is using computer output as a second opinion and making a final decision. It is obvious that higher performance of computer will lead to better quality on the final diagnosis, however the performance level of computer does not have to be equal to or higher than that of physicians [2]. Rather than that focusing on the synergistic effect obtained by combining ability of physicians and computers are the most important factor in CAD, and the current CAD has become a practical tool in many clinics with attractive benefits [2].

A new technology for breast cancer screening, MTI, was recently introduced as an alternative scheme to diagnose breast cancer that has many advantages over the

mammogram such as cost-efficient and requiring less processing-time. A study manifested that MTI surpasses over standard techniques, namely MRI, x-ray mammography, and ultrasound in aspects of low health risk, non-invasive, inexpensive, and minimal discomfort [5; 6]. MTI gauges dielectric properties of tissues, which are permittivity and conductivity, inside of the breast. Two parameters indicate the propagation of electromagnetic radiation that will be differ based on temperature, density, water content, and geographical conditions within breast mass. To our knowledge, instead of mammogram, only few research studies are ongoing using MTI as an input for the CAD system [7; 8; 9; 10]. The platform of input can have a flexibility by adopting artificial neural network in classification phase which is one of the phases in the CAD system. Unlike other classifiers such as Support Vector Machine (SVM), Artificial Neural Network has a few points to be cautious. The number of hyper-parameter will be different per network designer. Its combination can be quite various, and if it is being too many or few the network is hard to avoid overfitting or underfitting problem, respectively. Mathematically, it is same as trying to solve indeterminate equation. Moreover, while SVM guarantee a global optimum value, which is the lowest value of cost function, the ANN is possible to converge to local minimum value called local minimum problem. Our work focuses on avoiding above mentioned circumstance and to have reliable classification result, that of Matthews Correlation Coefficient (MCC), Recall, Specificity, Accuracy, and Precision.

In this paper, we propose an optimized model, MLP using DLR, in order to obtain better performance for binary classification that can be plugged into the CAD software platform. Figure 1 illustrates our model. In the preprocessing step, only 3 top ranked

features are selected by a correlation-based score in order to optimize the size of neural network. This is so the network does not have too many weights compared to small dataset; otherwise this might cause indeterminate equation problem. Then the feed forward MLP alters weights by using standard back-propagation with DLR in learning phase. DLR will compensate a higher learning rate for training points belonging to minority class, which optimizes model when dataset contains unbalanced class. The overall performance of this model outperforms the conventional model. Our model promises better performance than the existing model, hence practitioner can obtain more reliable result.

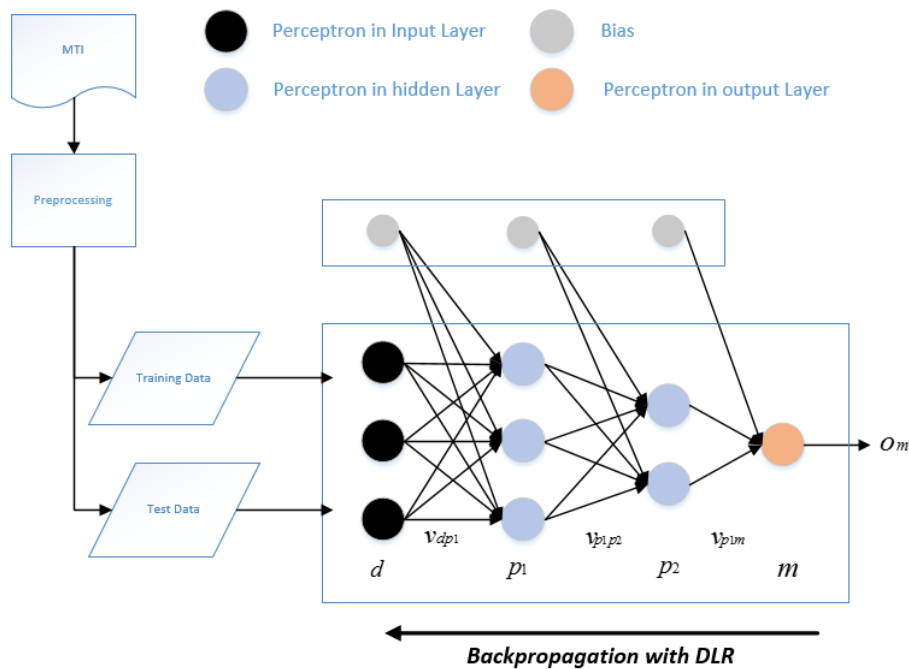


Figure 1. Diagram of proposed MLP model using DLR

The remainder of this thesis is organized as follows: Section 2 is a literature review that contains background of CAD system and MLP. The related work is mentioned in Section 3. Section 4 describes our proposed model. Section 5 shows some

experimental results and analysis. Finally, Section 6 concludes my thesis and gives some directions for future work.

## 2. LITERATURE REVIEW

### 2.1. COMPUTER AIDED DIAGNOSIS SYSTEM

The CAD system provides attractive benefits for image screening interpretation. Even for the expertized radiologists, interpreting screening mammogram is not an easy task because of its work load caused by the number of images and its quality [11]. To overcome this problem, using double rendering without increasing recall rate requires much more manpower, and accordingly it leads to increase cost. However, when it comes to the CAD, it can offer a cost-effective alternative to double rendering by acting as a second reader [11]. The way of utilizing the CAD system is that a radiologist at first rendering interprets first on the mammogram, and then compare it to the results from the computer at second rendering in order to check if he or she missed or unchecked what they did at the first rendering [11]. As described, the computer rather being a subject of making decision, is providing additional information to physicians in cost-effective way, that of requiring less manpower and time, with even higher diagnose quality [11]. Currently, the CAD system is widely being used in many clinics with mentioned advantages, and also developing new algorithms for the CAD system of breast cancer is active research field in academia [11]. Figure 2 shows how a CAD system can be utilized for early detection of breast cancer, which was one of the research studies I have done.



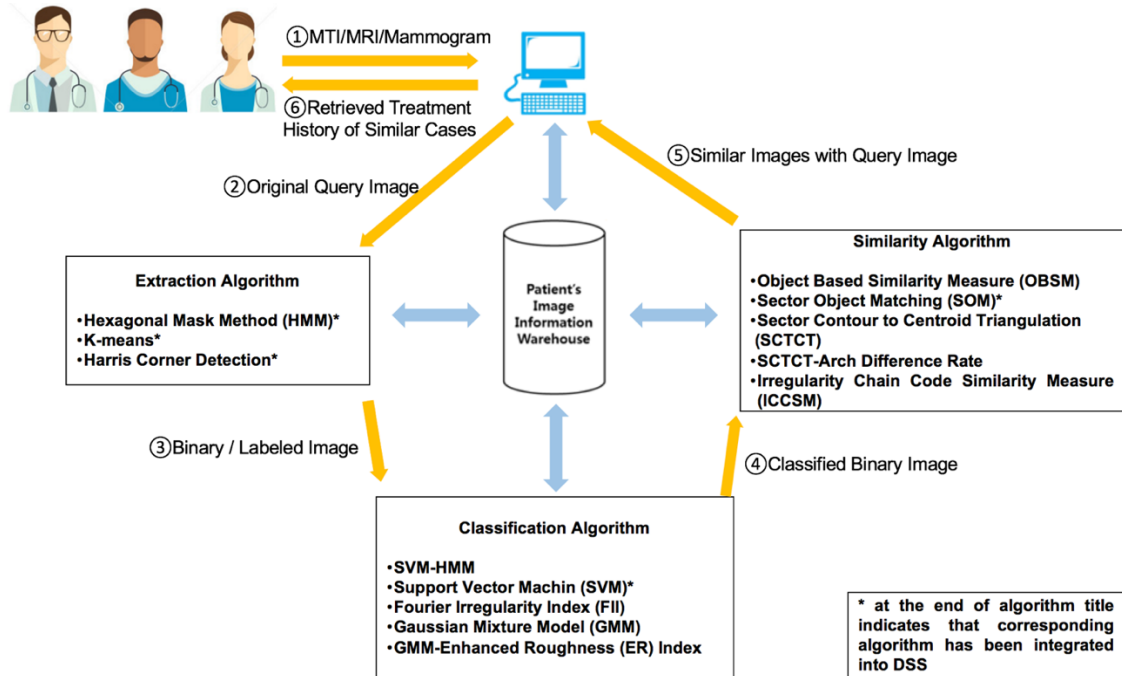


Figure 2. Expert system: Diagram of CAD system for early detection of breast cancer

Typically, the CAD system consists of multiple phases such as 1) preprocessing phase, 2) feature extraction & selection, and 3) classification. In preprocessing phase, given medical images are enhanced to have higher quality through several procedures such as removing noise and contrast enhancement. Various features are extracted from the processed image in feature extraction and selection phase. Domain experts and feature engineers choose features to be extracted based on heuristic manner, that of what kind of features will retrieve best performance in the classification phase. This step is able to be skipped depending on the kinds of classifier, for instance typical Convolutional Neural Network (CNN) does not require human being to pick particular feature set. Otherwise, having a reasonable strategy for selecting features is essential since it can be time-consuming task, the combination of selected feature set can be numerous, and it will directly affect the quality of classification. Finally, the classifier makes decision with

given feature set in the classification phase. Practically, since developer has lots of options to choose for the classifier such as Naïve Bayes Classifier, K-Means Clustering, SVM, or ANN, it is important to choose the most feasible one based on given problem and expected output.

## 2.2. MULTI-LAYER PERCEPTRONS

A MLP is a type of ANN such as Single Layer Perceptron (SLP) or Self-organizing Map (SOM) that is getting a lot of attention recently as a fundamental architecture of deep neural network. As its name states by itself, MLP comprises at least three layers: one input layer, one or more hidden layer, and one output layer. Each layer has many perceptron linked with another in the next layer, and the connections have their own value called weight. Eventually, the well-trained network should have the optimal weight set. What the hidden layers does is basically similar to what a kernel does in SVM that of projecting feature vectors to high dimensional space and finding hyper-plane separating training points properly. In MLP, a similar effect to projecting feature vectors to high dimensional space can be achieved through the multiple layers of perceptron. Each layer constructs a feature vector for given data based on the assigned weight value. The feature from the previous layer is passed to the next layer through the non-linear activation function. This results in a change in the feature space in which the data can be projected, and the next layer can also map the input to the new feature space.

A perceptron is a mathematical model of biological neuron. Given features in perceptron is transmitted in a similar way to propagation of electrochemical stimulation in neuron. Each perceptron takes weighted sum of the information from its nearby

perceptron like dendrite and pass it to activation function. Only filtered out information based on activation rule is passed to next perceptron which is called feed-forward computation. When it comes to the activation function, a non-linear function is adopted to determine which information is significant enough to be passed. Typically, a binary or bipolar sigmoid functions are the options for an activation function. Recently, various research studies on the CNN have proposed many other choices, such as Rectified Linear Unit (ReLU), which can tackle the gradient vanishing problem [12].

The ultimate objective of training in MLP is to find weight set that of minimizing result of cost function, ideally making it to be 0. In supervised learning, a cost function measures difference between prediction value from the model and actual expected output. For every iteration, all weights in the network will be updated in the way of reducing the output of the cost function. The mechanism of deciding up to what degree should be updated is called learning algorithm and gradient descent is dominant technique.

Due to the nature of the cost function, the next weight should be made equal to the learning rate in the opposite direction of the direction of the derivative of the cost function at the current weight. At this point, the classification results vary considerably depending on the learning rate. With the use of too small learning rate, the cost function requires much time and computation to reach minimum value. On the other hands, if the learning rate is too large, the cost function may not decrease on every epoch and it may not even converge. Worst part of ANN is that even though we happened to use a desirable learning rate, the network might not be able to reach the global minimum value, but reach the local minimum value, which is called minima problem.

### 3. RELATED WORK

#### 3.1. FEATURE EXTRACTION AND SELECTION

The strategy of feature extraction and selection is important because the prediction quality of a classifier depends on which subset of features are used as input. According to the work by Samaneh et. al [13], the permittivity value, one of the measurements obtained through MTI, depends on the water content that tissue contains.

Typically, tissue containing cancerous tissue has relatively high water content, while healthy tissues have lower water content. Therefore, in order to detect suspicious region that might have cancerous tissue, they divide data of sample into two groups based on the distribution of the permittivity. For this, they employed K-Nearest Neighbor. K-Nearest Neighbor is a geometric way to group objects with similar spatial distances of feature vectors into separate clusters. First, after dividing into cancerous tissue and healthy tissue in two groups by Euclidean Distance, then numerous subset of feature of each group were measured as follows:

- 1) Mean value: The mean calculates the average value of permittivity in different lesion. The lesion that contains tumor has a higher mean than those of normal tissue.
- 2) Maximum and minimum value of permittivity in probable tumor area: These values indicate the domain of cancerous tissues permittivity.
- 3) Entropy: Entropy measures the amount of disorder in permittivity data.  
Entropy calculated from the permittivity as per the following Equation (1).

$$Entropy = \frac{1}{n} \sum_{i=1}^n per_i * \log(per_i) \quad (1)$$

- 4) Energy: Energy represents the orderliness of permittivity data. Energy is generally given by the mean squared value of a signal. Energy is calculated from permittivity as per the following Equation (2).

$$Energy = \frac{1}{n} \sum_{i=1}^n (per_i)^2 \quad (2)$$

### 3.2. CLASSIFICATION

Samaneh et. al [13] adopts MLP, one of the various ANN type, as a classifier. They used a network with a total of four layers, with two hidden layers each having 30 neurons. The input layer uses the previously obtained feature vectors and the output layer is implemented as a binary classifier with benign or malignant. The training process is supervised-learning using Levenberg-Marquardt back propagation to minimize the cost-function. Here, the cost-function uses the sums-of-square error function as below:

$$E = \frac{1}{2} \|e\|^2 \quad (3)$$

Where e is the difference between networks predicted output and the expected value.

The conditions that cause the training to be terminated use the default values set in *trainlm*, a function given in MATLAB. They performed testing using the weight of the trained network through the above mentioned configuration.

#### 4. PROPOSED MODEL

Originally, MLP does not require a human being to choose particular feature since when massive dataset is given, machine can extract correlated features by itself during learning phase. We have few datasets, however, specifying some features as an input for MLP was compulsory. Along that fact, after separating the patient's dataset into 2 groups of cancerous and healthy tissue by using K-Nearest Neighbor algorithm, we extracted set of commonly used intensity and texture based features per each group as following [14]:

- 1) Mean value
- 2) Maximum and minimum value of permittivity/conductivity in probable healthy/tumorous area
- 3) Entropy
- 4) Energy
- 5) Skewness: The skewness measures asymmetry of the probability distribution of a real-valued random variable about its mean.

$$Skewness = \frac{E(x - \mu)^3}{\sigma^3} \quad (4)$$

- 6) Kurtosis: The kurtosis measures tailedness of the probability distribution of a real-valued random variable.

$$Kurtosis = \frac{E(x - \mu)^4}{\sigma^4} \quad (5)$$

- 7) Ratio of Healthy/Tumorous Tissue within breast mass

As a result, we end up with 52 different features total, which is too large number as input vector. A more detailed explanation is described in the Section 4.1, but we cannot use all the features as an input vector for the network because the number of input

is related with the size of the network and the size of network does matter to its performance. We thus selected only 3 features listed below ranked on the top among 52 features based on the correlation-based score, which is shown in Equation (6) [15; 16].

- 1) Ratio of health tissue within breast mass
- 2) Ratio of tumor tissue within breast mass
- 3) Maximum value of conductivity in healthy tissue

$$Score = \frac{\sum_{i=1}^k corr(c, f_i)}{\sqrt{k + 2 * \sum_{i=1}^{k-1} \sum_{j=1}^{k-1} corr(f_i, f_j)}} \quad (6)$$

#### 4.1. CLASSIFICATION PHASE

Selected features will be used as an input into the MLP that consists of 1 input layer, 2 hidden layers (each layer contains 3 and 2 perceptron respectively), and 1 output layer. How we came up with this setting is based on following fact. First of all, it has been proved that using three-layer perceptron may be able to construct complex decision regions faster with back-propagation than two-layer perceptron [17]. Second, the network size is determined by following equation.

$$(d + 1) * p + (p + 1) * m \quad (7)$$

where  $d$  is the number of features that goes into the input layer,  $p$  is the number of perceptron in hidden layer, and  $m$  is the number of perceptron in output layer. When size of network is larger than the number of samples, then it might cause a low performance and lack of reliability for network result. For instance, training network that consists of 100 inputs, 30 perceptron in hidden layer, and 10 perceptron in output layer with 100 samples is same as trying to solve equation that has 3340 variables, which is weights in

network, with 100 value and it is an indeterminate equation. To avoid referred problem, our model has been optimized with the setting as mentioned above.

Our model works as follows. Once optimized size of network takes some features coming from preprocess phase, feed-forward MLP will yield an output, which is a prediction value, and measures the discrepancy between the output yielded by neural network and actual value that “correct” or desired output of corresponding sample. Suppose that we have a set of desired output  $T = (t_1, t_2, \dots, t_p)$  and a set of output yield by network  $O = (o_1, o_2, \dots, o_p)$ . Then the discrepancy will be produced by following formula.

$$E = \frac{1}{2} \sum_{i=1}^p T_i - O_i \quad (8)$$

Next, the weight between perceptron will be updated through the learning process, which is also known as error-correction process. The loss function described by Equation (8), leads to a learning rule commonly known as delta rule [18]. The next weight set is determined by adding particular amount of value, say  $\Delta v$ , to current the weight set. The value  $\Delta v$  is determined by Equation (9) that is derived from the Equation (8). As a result, the learning process described above can be put succinctly as below equation.

$$v(h + 1) = v(h) + \Delta v = v(h) - \rho \frac{\partial E}{\partial v} \quad (9)$$

where  $h$  is epoch and  $\rho$  is learning rate.

In the Equation (9), all the factors, except the value of  $\rho$ , are already determined. The value of  $\frac{\partial E}{\partial v}$  indicates the gradient of the graph in Figure 3 at point of current weight set. For example, in Figure 3, suppose that the network is initialized to have a value of



weight set located just to the right of point A. Then the gradient is determined to be negative value. The sign of the gradient value tells what direction the next weight update should take. Thus, the remaining part for us to decide is determining to what magnitude the next weight update should take. This is where the learning rate comes into play. Too small learning rate makes system slow that is obviously not good for performance and too large learning rate causes neural network to be dispersed from the desired minimum.

Figure 3 illustrates that a network with an improper learning rate can be a problematic.

The path from point *A* to *C* shows that if the network has a small learning rate, the global optimum can be reached, but the processing time can be too long. On the other hand, as the path from point *A'* to *B* shows, when the network has a large learning rate, it fails to converge to the global optimum and falls into the local minimum.

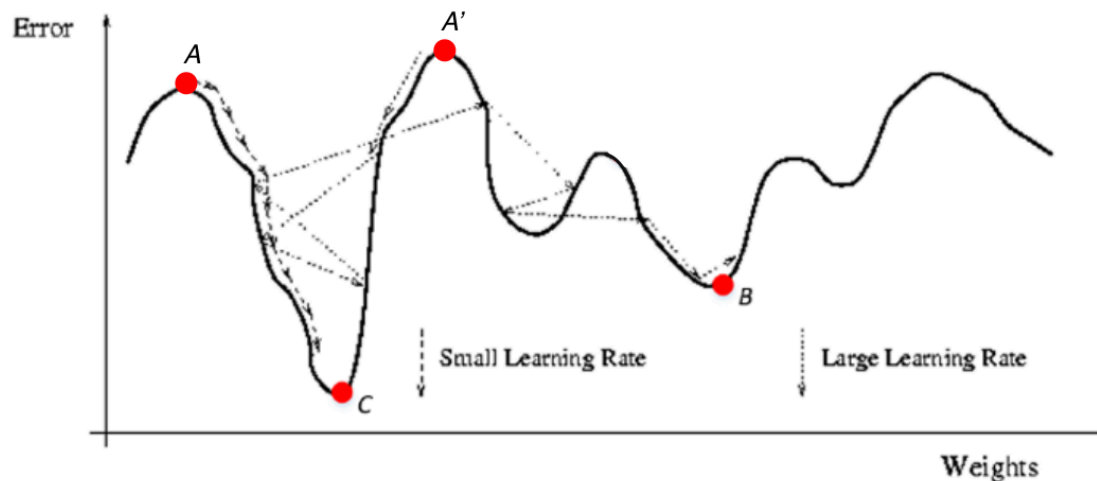


Figure 3. The influence of learning rate on the weight changes

So, determining proper learning rate is one of the most important factor for neural network. It has been proved that fixed learning rate will cause a low efficiency [19]. Our network will be more fragile on this concern because of lack of data and unbalance on class. To overcome this issue, we designed a logic to assign a learning rate dynamically

to each training points. Main idea is simple as shown in algorithm<sup>1</sup> 1. During the learning phase, looking at the accumulated trained samples and when particular class overwhelms the other class, then assigning a high learning rate onto training point belonging to minority class and a low learning rate onto training point belonging to majority class dynamically. By doing so, network can assign more reasonable weights between perceptron in the way that having network not to be overfitted with training data belonging to majority class by compensating training data belonging to minority class with high learning rate. Doing so neural network trained by sample having unbalance on class can be sensitive on minority class so that network can avoid possible local minima.

During the learning phase, neural network will keep updating the weight between perceptron in the way of reducing the error,  $E$ , shown in the Equation (8). Ordinarily, there are two different strategies for updating weights between perceptron in neural network. One is batch mode that updates  $\Delta v$  all at once with its average value after training all samples. The other one is pattern mode, which is used in our work that updates  $\Delta v$  right after training each sample. It is obvious that pattern mode is more sensitive on noise or newly provided samples and practically it shows better performance than batch mode [20].

---

<sup>1</sup> Available at [https://github.com/chulwoopack/MLP\\_using\\_DLR](https://github.com/chulwoopack/MLP_using_DLR)

```

1  %(flag is TRUE when system trained Cancer case more than Normal case)
2  if flag then
3      % Cancer excesses Normal
4      if current Image is Cancer
5          % Low Learning Rate
6          learning rate = # of normal cases / total
7      else
8          % High Learning Rate
9          learning rate = # of cancer cases / total
10     else
11         % Normal excesses Cancer
12         if current Image is Cancer
13             % High Learning Rate
14             learning rate = # of normal cases / total
15         else
16             % Low Learning Rate
17             learning rate = # of cancer cases / total
18     end

```

Algorithm 1. Dynamic Learning Rate

The last step is determining when machine is to be terminated. Among several methods, for example, using epoch, using Mean Squared Error (MSE), or using validation set, we combined MSE and epoch as a condition to terminate the neural network. Since no general condition does exist for deciding whether network has converged or not, adjusting conditions based on given network size and couple of experiments are required. For instance, if network goes through too long epoch or too small MSE value, it will make network to be overfitted that could lead neural network cannot classify test sample correctly. So we set the neural network to be terminated when epoch is less than 1000 and MSE is less than 0.1 based on experimental results.

$$epoch < 1000 \text{ and } MSE = \frac{\sum_{i=1}^N E_i}{N} < 0.1 \quad (10)$$

where N is the number of sample.

## 5. RESULT AND ANALYSIS

### 5.1. EVALUATION PHASE

To evaluate our model, we decided to measure precision, sensitivity, specificity, accuracy, and MCC with confusion matrix that was used in [6] for comparison purpose.

- Precision: The proportion of the true positive against all the positive results.

$$\text{Precision} = \frac{TP}{TP + FP}$$

- Sensitivity or Recall: Ability of test to identify positive result; correctly identify patient has cancer who has cancer.

$$\text{Sensitivity or Recall} = \frac{TP}{TP + FN}$$

- Specificity: Ability of test to identify negative result; correctly identify patient is healthy who has no cancer.

$$\text{Specificity} = \frac{TN}{TN + FP}$$

- MCC: Quality of binary classifications; a value between -1 to +1. A coefficient of +1 indicates a perfect prediction, 0 no better than random prediction, and -1 indicates total failure between prediction and observation.

$$\text{MCC} = \frac{(TP * TN - FP * FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

where true positive ( $TP$ ) is the number of samples that correctly identified as cancer, true negative ( $TN$ ) is the number of samples that correctly identified as normal,

false positive (*FP*) is the number of samples that incorrectly identified as cancer, and false negative (*FN*) is the number of samples that incorrectly identified as normal.

## 5.2. RESULTS

In this work we used the same dataset and data shuffling scheme as used in [6] for comparison purpose, which is a set of data consisting of 30 breast MTI results that contains permittivity, conductivity, and coordinates of each tissue within breast mass from clinic trial at Seoul National University, Korea.

Table 1 and Table 2 show the confusion matrix of proposed model and conventional model respectively. These tables describe the average value of TP, TN, FP, and FN from results of 15 datasets. The average value of precision, recall, specificity, and accuracy for 15 datasets are 90.9%, 66.67%, 97.14%, and 88.0% respectively by proposed model.

Table 1. Confusion matrix of MLP using DLR

		Actual Condition determined by Doctor		
		Positive	Negative	
Test Outcome	Positive	2 ( <i>TP</i> )	0.2 ( <i>FP</i> )	Precision = 90.9
	Negative	6.8 ( <i>FN</i> )	1 ( <i>TN</i> )	
		Recall = 66.67	Specificity = 97.14	Accuracy = 88.0

Table 2. Confusion matrix of conventional model

		Actual Condition determined by Doctor		
		Positive	Negative	
Test Outcome	Positive	1.9 ( <i>TP</i> )	0.6 ( <i>FP</i> )	Precision = 81.78
	Negative	6.4 ( <i>FN</i> )	1.1 ( <i>TN</i> )	
		Recall = 63.33	Specificity = 85.54	Accuracy = 82.67

Overall measures outperform conventional model especially on specificity as shown in Figure 4.

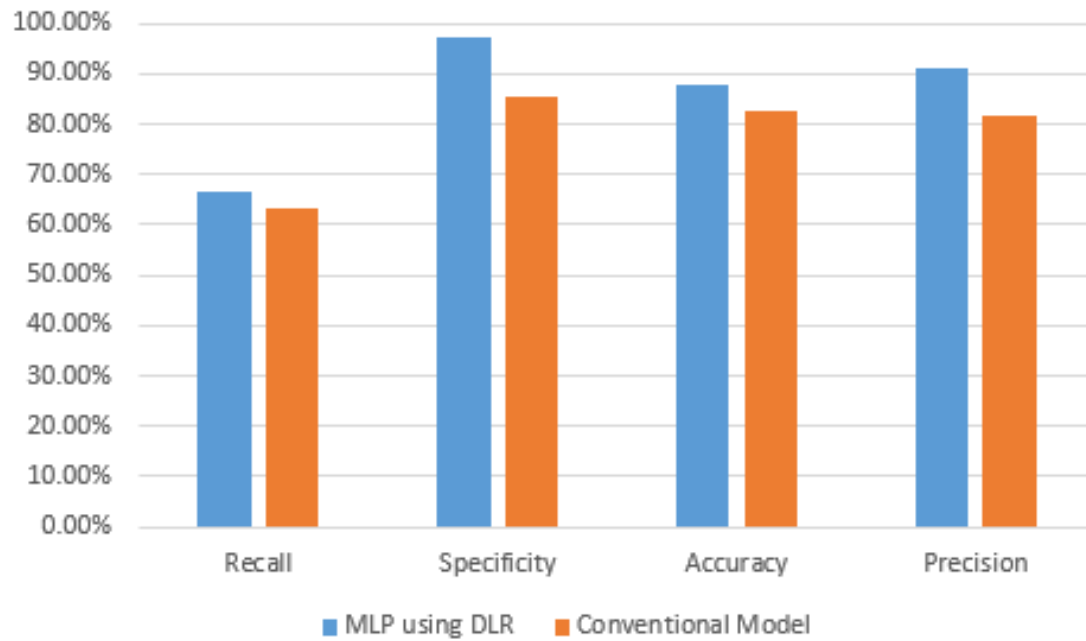


Figure 4. Average value of confusion matrix analysis

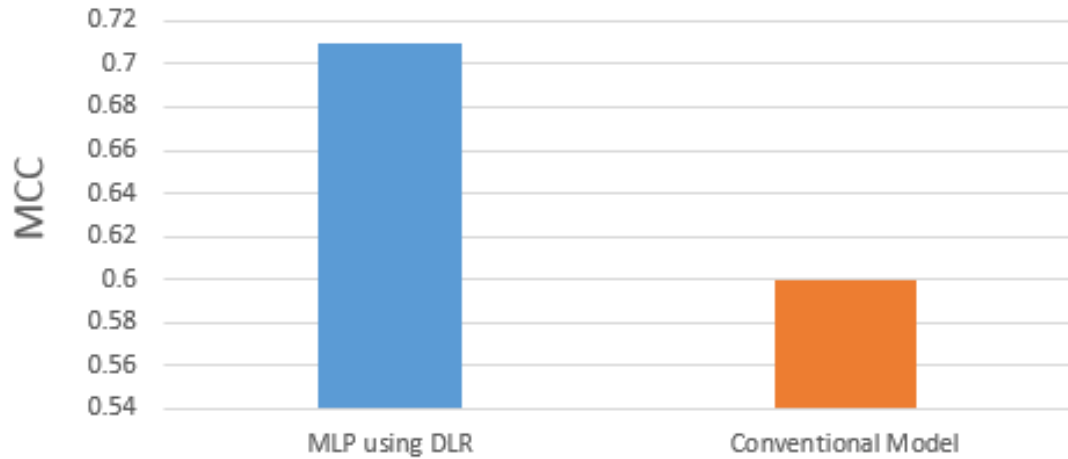


Figure 5. MCC value comparison

Proposed model also shows a better quality on classification in terms of MCC value. Our model produces a 0.71 while the conventional model produces a 0.6 as shown in Figure 5. That implies that proposed model promises more reliable outcome on binary classification with stronger positive relationship.

Based on shown tables and figures, experimental results can be summarized optimizing size of neural network and assigning learning rate dynamically by earmarking higher learning rate onto each training data points of minority class can produce better performance. Also patient can anticipate saving cost from unnecessary biopsy with high specificity value, 97.14%.

## 6. CONCLUSION

A CAD software is still widely being developed to support practitioner as a second opinion, which is heavily affected by classification performance [3; 4]. Though novel technology; which is MTI that showing various advantages over standard techniques in aspect of low health risk, non-invasive, inexpensive, and minimal discomfort; was released, corresponding classification tool that showing outstanding performance has been hardly suggested [5; 6]. In this paper, we proposed MLP using DLR, which is an improved version of a model suggested by Samaneh et al., in order to optimize classification phase that will be plugged into the CAD software platform with promising robust performance [13]. Since the existing model suggested by Samaneh et al. has excessive number of perceptron in hidden layer and uses static learning rate with small amount of dataset having unbalance on class, it might produce unreliable result caused by either indeterminate equation problem or overfitting problem. Comparing to these concerns, in our model, optimized size of neural network guarantees the learning process not to fall into indeterminate equation problem by not having excessive number of weights between each other perceptron in neural network compared to number of sample, so that it can produce a reliable result. Also our model uses DLR during learning phase to dynamically assign learning rate onto each training point based on which class overwhelms the other class. Assigning higher learning rate onto a training point belonging to minority class makes neural network possible to escape from local minima, which is typical jeopardy of ANN. This proposed classification model can optimize the CAD software platform by being plugged into it. The results show that our model outperforms existing model not only for the performance such as recall, specificity,



accuracy, and precision, but also for the quality, and thus it empowers physicians to make better decision on breast cancer screening in early stage, as it also alleviates the cost burden from the patients.

As future work, we aim to focus on designing deep layer neural network without any feature selection when massive data is given. Then we intend to deal with optimizing classification phase to construct robust CAD system for breast cancer screening.

## LITERATURE CITED

1. "Breast cancer: prevention and control." World Health Organization. Accessed July 08, 2017. <http://www.who.int/cancer/detection/breastcancer/en/>.
2. Doi, Kunio. "Computer-aided diagnosis in medical imaging: historical review, current status and future potential." *Computerized Medical Imaging and Graphics* 31, no. 4 (2007): 198-211.
3. Baker, Jay A., Eric L. Rosen, Joseph Y. Lo, Edgardo I. Gimenez, Ruth Walsh, and Mary Scott Soo. "Computer-aided detection (CAD) in screening mammography: sensitivity of commercial CAD systems for detecting architectural distortion." *American Journal of Roentgenology* 181, no. 4 (2003): 1083-1088.
4. Sharma, Shubhi, and Pritee Khanna. "Computer-aided diagnosis of malignant mammograms using Zernike moments and SVM." *Journal of Digital Imaging* 28, no. 1 (2015): 77-90.
5. Santorelli, Adam, Emily Porter, Evgeny Kirshin, Yi Jun Liu, and Milica Popovic. "Investigation of classifiers for tumor detection with an experimental time-domain breast screening system." *Progress In Electromagnetics Research* 144 (2014): 45-57.
6. Noghianian, S. "Microwave Tomography for Biomedical Quantitative Imaging." *J Elec Electron* 1, no. 3 (2012).
7. Floyd, Carey E., Joseph Y. Lo, A. Joon Yun, Daniel C. Sullivan, and Phyllis J. Kornguth. "Prediction of breast cancer malignancy using an artificial neural network." *Cancer* 74, no. 11 (1994): 2944-2948.
8. Wu, Yuzheng, Maryellen L. Giger, Kunio Doi, Carl J. Vyborny, Robert A. Schmidt, and Charles E. Metz. "Artificial neural networks in mammography: application to

- decision making in the diagnosis of breast cancer." *Radiology* 187, no. 1 (1993): 81-87.
9. Zhang, Wei, Maryellen L. Giger, Yuzheng Wu, Robert M. Nishikawa, and Robert A. Schmidt. "Computerized detection of clustered microcalcifications in digital mammograms using a shift- invariant artificial neural network." *Medical Physics* 21, no. 4 (1994): 517-524.
  10. Christoyianni, I., A. Koutras, E. Dermatas, and G. Kokkinakis. "Computer aided diagnosis of breast cancer in digitized mammograms." *Computerized Medical Imaging and Graphics* 26, no. 5 (2002): 309-319.
  11. Rangayyan, Rangaraj M., Fabio J. Ayres, and JE Leo Desautels. "A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs." *Journal of the Franklin Institute* 344, no. 3 (2007): 312-348.
  12. Havaei, Mohammad, Axel Davy, David Warde-Farley, Antoine Biard, Aaron Courville, Yoshua Bengio, Chris Pal, Pierre-Marc Jodoin, and Hugo Larochelle. "Brain tumor segmentation with deep neural networks." *Medical Image Analysis* 35 (2017): 18-31.
  13. Aminikhanghahi, Samaneh, Wei Wang, Sung Shin, Seong H. Son, and Soon I. Jeon. "Effective tumor feature extraction for smart phone based microwave tomography breast cancer screening." In *Proceedings of the 29th Annual ACM Symposium on Applied Computing*, pp. 674-679. ACM, 2014.
  14. Sachdeva, Jainy, Vinod Kumar, Indra Gupta, Niranjana Khandelwal, and Chirag Kamal Ahuja. "Segmentation, feature extraction, and multiclass brain tumor classification." *Journal of Digital Imaging* 26, no. 6 (2013): 1141-1150.

15. Fear, Elise C., Paul M. Meaney, and Maria A. Stuchly. "Microwaves for breast cancer detection?." *IEEE Potentials* 22, no. 1 (2003): 12-18.
16. Liu, Huiqing, Jinyan Li, and Limsoon Wong. "A comparative study on feature selection and classification methods using gene expression profiles and proteomic patterns." *Genome Informatics* 13 (2002): 51-60.
17. Anderson, Dana Z., ed. *Neural Information Processing Systems: Proceedings of a conference held in Denver, Colorado, November 1987*. Springer Science & Business Media, 1988.
18. Kubat, Miroslav. "Neural networks: a comprehensive foundation by Simon Haykin, Macmillan, 1994, ISBN 0-02-352781-7.-." *The Knowledge Engineering Review* 13, no. 4 (1999): 409-412.
19. Yu, Xiao-Hu, Guo-An Chen, and Shi-Xin Cheng. "Dynamic learning rate optimization of the backpropagation algorithm." *IEEE Transactions on Neural Networks* 6, no. 3 (1995): 669-677.
20. LeCun, Y., L. Bottou, and G. Orr. "Efficient BackProp in Neural Networks: Tricks of the Trade (Orr, G. and Müller, K., eds.)." *Lecture Notes in Computer Science* 1524, 9-48.