Breast Cancer Classification Using Artificial Neural Networks

Aditya Jain
Dept. of Computer Science and Engineering
National Institute of Technology, Calicut
M200567CS

Aditya Semwal
Dept. of Computer Science and Engineering
National Institute of Technology, Calicut
M200599CS

***Abstract* — The purpose of this project is to use deep learning methodology to predict nature of breast cancer tumor. This paper proposes design of Artificial Neural Network to tackle the chosenproblem,moreover it also aims to compare performance of various optimizers and loss functions used in designing of Artificial Neural Network. We proposed that model with Convolution layers, having Drop - Out and Batch - Normalization, with slow learning rate is best choice since it provide highest accuracy on validation data.**

Keywords: Bioinformatics, Deep Learning, ANN Data Processing, DataMining, Machine Learning.

I. Introduction

Breast cancer refers to cancer from a malignant tumor in the cells of the breast tissue [1]. A malignant tumor is a group of cancer cells that can grow into surrounding tissues or spread to distant areas of the body. Breast cancer is an uncontrolled multiplication of cells in breast tissue. A group of rapidly dividing cells may form a lump or architectural distortions. The second leading cause of death among women is breast cancer, as it comes directly after lung cancer. Breast cancer is a life-taking disease and early detection can certainly reduce the rate of mortality.

Most of the types of breast cancers are easily diagnose by Tissue Biopsy [1]. In which examination of affected tissue is done by experts and nature of tissue is identified, we can automate this process by feeding data generated from biopsy to a computer and using various machine learning method to identify nature of tumor.

II. Problem Definition

Problem can be defined as given a set of parameters describing breast cancer tissue information, design an Artificial Neural Network [2] to classify tissue as either *Malignant* or *Benign*.

We used *Wisconsin Breast Cancer Dataset* [3] as input of the model. This is one of the well known dataset used by various researcher in past to carry out various research in machine learning field. This dataset can be obtained from UCI machine learning repository [3], which is well known resource for data mining and machine learning enthusiasts to get various kind of data.

Wisconsin Breast Cancer Dataset contains 569 rows each having 32 attributes, this features are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. They describe characteristics of the cell nuclei present in the image.

This dataset combined with various machine learning and deep learning techniques is then used to train model which will be capable of predicting nature of tissue once model is supplied with previously not seen information.

III. Literature Survey

Following survey describes Artificial Neural Network which will be useful in implementing the project. It also provides an insight for previously done work by researchers in detecting Breast Cancer using similar or different techniques.

A. Artificial Neural Network

Idea of Artificial Neural Network is based upon biological neuron. A neuron (or nerve cell) is a

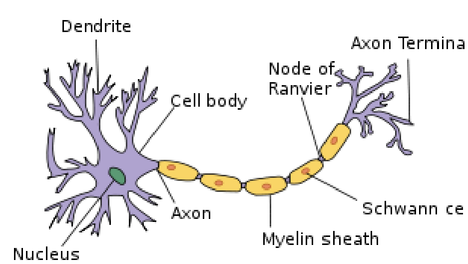

Fig. 1. Biological Neuron

special biological cell that processes information.

we use artificial neuron for training of ANN's, and train them as:

Supervised Learning, in which model is trained on labelled dataset, a labelled dataset is one which have both input and output parameters. Supervised learning can be used to perform tasks like pattern recognition and regression.

Unsupervised Learning is the training of model using information that is neither classified nor labelled. Unsupervised learning is used in estimation problems, such as clustering, the estimation of statistical distributions, compression and filtering.

Reinforcement learning , in this model/agent take action/learn to take action in environment which maximize the rewards. This technique is used in vehicle routing, video games, natural resource management and medicine.

B. Related Works

Breast Cancer Classification has been interest of many researchers over the years. We have studied recent development in this area, following is glance of the related work that has been done by many researcher over past decade.

In [10], H.D. Cheng et. al. uses Ultrasound images of breast cancer tissue, image processing and machine learning techniques to automatically classify breast cancer tissue into benign and malignant.

In [11], Mehrbakhsh Nilashi et. al., develop a new knowledge-based system for classification of breast cancer disease using clustering, noise removal, and classification techniques. They used Expectation Maximization (EM) as a clustering method to cluster the data in similar groups. Then they use Classification and Regression Trees (CART) to generate the fuzzy rules to be used for the classification of breast cancer disease in the knowledge-based system of fuzzy rule-based reasoning method. To overcome the multi-collinearity issue, they also incorporate Principal Component Analysis (PCA) in the proposed knowledge-based system.

In [12], authors Fabio Alexandre Spanhol et. al., uses a publicly available dataset of breast cancer histopathological images from BreaKH and Artificial Neural Network to tackle the problem of breast cancer classification. Authors used various deep lerning techinques to train ANN and perform classification.

In [13], authors M. Saritas and A. Yasar, uses breast cancer classification as metric to measure the performance of ANN and Naive Bayes Classification algorithm.

In [14], S. W. Lam et. al. uses proteomic technologies to identify protein and it's subtype to classify breast cancer into different subtypes. They also discuss important aspects of the potential usefulness of proteomics for discovery of breast cancer-associated protein biomarkers in the clinic.

In [15], Abdel-Zaher et. al., developed Computer Aided Diagnosis System, using deep belief network unsupervised path followed by back propagation supervised path. They also use Wisconsin Breast Cancer Dataset to perform classification. They also propose two phase method for Deep Belief network, which gives more accurate result than single pas method.

In [16], Marcano-Cedeño et. al., propose training algorithm is inspired by the biological metaplasticity property of neurons and Shannon's information theory. They apply this algorithm on ANN and use Wisconsin Breast Cancer Dataset to perform classification of breast cancer and to measure performance of algorithm.

This conclude literature survey, there are many related work done in this area which is very similar to what we have presented here.

IV. Proposed Method

A multi phase method is used to build and train model for classification. First phase is data processing, second phase is building the model and third phase is training the model.

A. Data Processing

Main goal of this process is to make data ready for various machine learning algorithm. Work done in this phase will be removing NULL data from data set, performing normalization, standardization, feature encoding and scaling of data. This all different kind of data processing technique will help us to improve our model efficiency.

In Data Processing phase, to perform standardization of data *StandardScaler* utility is used, it used following formula to perform standardization.

$$new\ value = \frac{current\ value - mean}{standard\ deviation}$$

B. Building Neural Network Model

In this phase *"Keras: The Python Deep Learning library"* [5] was used to build model for Artificial Neural Network. Keras is high level API which allows to create Artificial neural network and complex connection between them in easy way.

We have created 4 different types of models,

- ANN - SLP (Single Layer Perceptron Network)

  It has only input and output layer. Input layer has 32 neurons and output layer has one neuron. See Fig. 4. for summary.

```
Model: "sequential_131"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense_295 (Dense)            (None, 32)                992
_________________________________________________________________
dense_296 (Dense)            (None, 1)                 33
=================================================================
Total params: 1,025
Trainable params: 1,025
Non-trainable params: 0
_________________________________________________________________
```

Fig. 4. ANN - SLP (Single Layer Perceptron Network)

- ANN - MLP (Multi Layer Perceptron Network)

  It has three layers, input layer having 32 neurons, one hidden layer with 64 neurons and output layer with single neuron. See Fig. 5. for summary.

- ANN - MLP with Convolution Layer (w/o Drop-Out and Batch-Normalization)

  It has two convolution layers, first one having 16 filters and second one having 32 filters. After the convolution layers, it has one flatten layer and two Dense layer, first Dense layer with 64 neurons and another with 1 neuron. See Fig.6. for summary.

```
Model: "sequential_132"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense_297 (Dense)            (None, 32)                992
_________________________________________________________________
dense_298 (Dense)            (None, 64)                2112
_________________________________________________________________
dense_299 (Dense)            (None, 1)                 65
=================================================================
Total params: 3,169
Trainable params: 3,169
Non-trainable params: 0
_________________________________________________________________
```

Fig. 5. ANN - MLP (Multi Layer Perceptron Network)

```
Model: "sequential_133"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv1d_162 (Conv1D)          (None, 29, 16)            48
_________________________________________________________________
conv1d_163 (Conv1D)          (None, 28, 32)            1056
_________________________________________________________________
flatten_79 (Flatten)         (None, 896)               0
_________________________________________________________________
dense_300 (Dense)            (None, 64)                57408
_________________________________________________________________
dense_301 (Dense)            (None, 1)                 65
=================================================================
Total params: 58,577
Trainable params: 58,577
Non-trainable params: 0
_________________________________________________________________
```

Fig. 6. ANN - MLP with Convolution Layer (w/o Drop-Out and Batch-Normalization)

- ANN - MLP with Convolution Layer (w/ Drop-Out and Batch-Normalization)

  It has two convolution layers, also additionally it has Drop-Out and Batch-Normalization layers also. See Fig. 7. for summary.
  Drop-Out layer randomly drops neurons from neural network, these dropped neurons are not considered in weight update of back-propagation algorithm, i.e. their weights are not updated.
  Batch-Normalization performs normalization of data and then transfers it to next layer.
  Both Drop-out and Batch-Normalization are used to avoid problem of Over fitting while training of Neural Network.

  In all the model output layer uses 'sigmoid'[18] as activation function, while all the others layers uses 'relu'[19] as activation function.

```
Model: "sequential_134"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv1d_164 (Conv1D)          (None, 29, 16)            48
_________________________________________________________________
batch_normalization_125 (Bat (None, 29, 16)            64
_________________________________________________________________
dropout_149 (Dropout)        (None, 29, 16)            0
_________________________________________________________________
conv1d_165 (Conv1D)          (None, 28, 32)            1056
_________________________________________________________________
batch_normalization_126 (Bat (None, 28, 32)            128
_________________________________________________________________
dropout_150 (Dropout)        (None, 28, 32)            0
_________________________________________________________________
flatten_80 (Flatten)         (None, 896)               0
_________________________________________________________________
dense_302 (Dense)            (None, 64)                57408
_________________________________________________________________
dense_303 (Dense)            (None, 1)                 65
=================================================================
Total params: 58,769
Trainable params: 58,673
Non-trainable params: 96
_________________________________________________________________
```

Fig. 7. ANN - MLP with Convolution Layer (w/ DropOut and BatchNormalization)

C. Training Model

Created model are trained using various optimizer and loss function.[17] For ANN- SLP and ANN - MLP various combinations of optimizer and loss function were used. Optimizers used are Adam, rmsprop, sgd. Loss functions used are mean squared error and binary cross-entropy. This is done to compare performance of various optimizer and loss function on breast cancer dataset.

ANN - MLP with Convolution layers are trained using Adam [17] Optimizer and Mean squared error [20] as loss function.

V. Results

Once ANN-SLP was trained, we are getting around 96% accuracy on validation data. Fig. 9-13 shows training process of ANN-SLP on different optimizer and loss function. From Fig. 9-13 we can see that, best choice of optimizer and loss function is sgd and mean square error as it provide highest accuracy and lowest loss, more over combination of Adam and mean square error, rmsprop and mean square error, sgd and Binary cross entropy are also feasible candidate for optimizer and loss function as seen from Fig. 8,10, 13. But Fig. 9 and Fig. 11 shows our model is over fitting on data as validation loss is more than training loss, so using Adam as optimizer and binary cross entropy as loss function on our data is not good choice, similarly rmsprop and binary cross entropy is also not good choice for breast cancer data set.

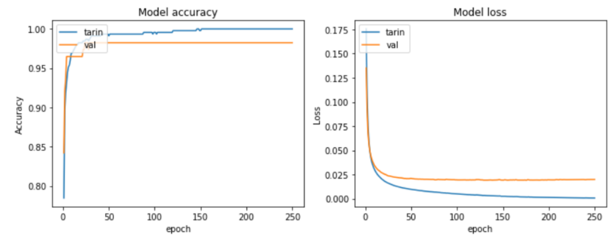

Fig. 8. ANN - SLP (Optimizer: Adam, Loss Function: Mean Squared Error)

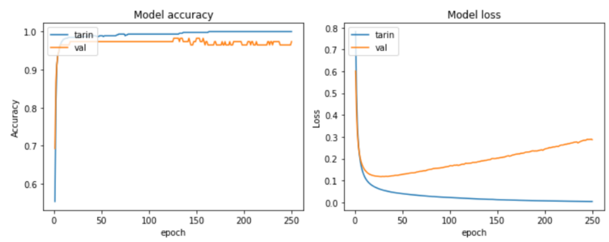

Fig. 9. ANN - SLP (Optimizer: Adam, Loss Function: Binary Cross-Entropy)

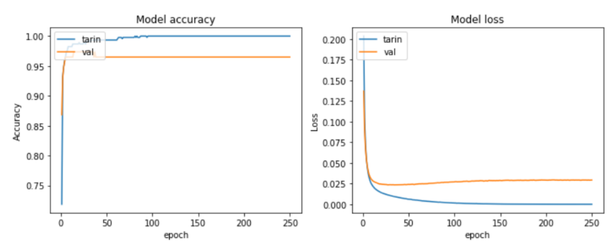

Fig. 10. ANN - SLP (Optimizer: rmsprop, Loss Function: Mean Squared Error)

When ANN-MLP is trained it gives slightly better performance than ANN-SLP, it has around 97% accuracy, in some cases it give around 98-99% accuracy. Fig. 14-19 shows learning process of ANN-MLP. We can see that learning curve of ANN-MLP is much smoother than ANN-SLP. Similar to ANN-SLP, ANN-MLP performs better when optimizer is 'sgd' and loss function is 'mean-squared-error'. Moreover combination of 'adam' and 'mean squared error', 'sgd' and 'binary cross entropy' also gives feasible accuracy as seen from Fig. 14, 19.

We can see that ANN-MLP is over-fitting, for optimizer and loss function combination are 'adam'

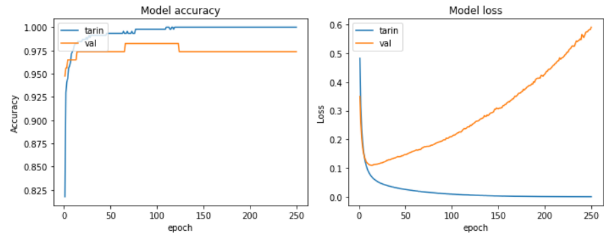

Fig. 11. ANN - SLP (Optimizer: rmsprop, Loss Function: Binary Cross-Entropy)

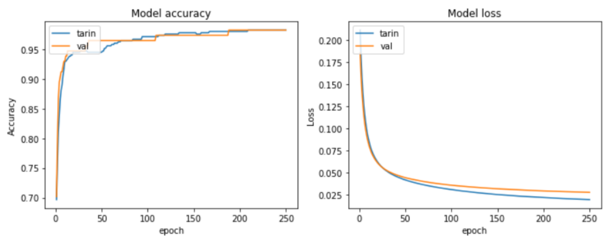

Fig. 12. ANN - SLP (Optimizer: sgd, Loss Function: Mean Squared Error)

and 'binary cross entropy', 'rmsprop' and 'mean square error', 'rmsprop' and 'binary cross entropy' as seen from Fig. 15, 16, 17.

When ANN-MLP with Convolution layer and without Drop-out and Batch-Normalization model was trained it gives around 97% validation accuracy. Even though it provide such high accuracy, we can see from Fig. 20 that there high variation in accuracy, more over validation loss is increasing which shows that our model is over-fitting.

This suggest the use of Drop-Out and Batch-Normalization. When model with Drop-Out and Batch-Normalization was trained, we can see that we are getting around 98% accuracy and model is also not over fitting which can be seen from Fig. 21. But there is lot of fluctuation while model is learning, this happens when we trained model with high learning rate. To avoid this same model was trained with learning rate of 0.0001. This does improve accuracy, now model accuracy is around 99% on validation data and learning curve is also smooth, whihc can be seen from Fig. 22.

VI. CONCLUSIONS

For breast cancer data set ANN-SLP and ANN-MLP works better when optimizer is 'sgd' and loss function is 'mean squared error'.

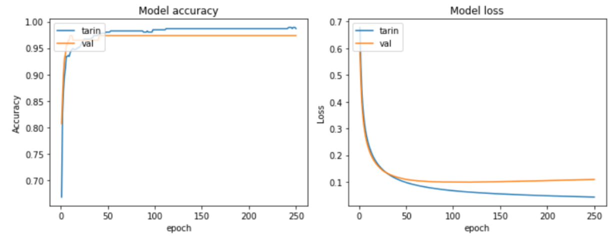

Fig. 13. ANN - SLP (Optimizer: sgd, Loss Function: Binary Cross-Entropy)

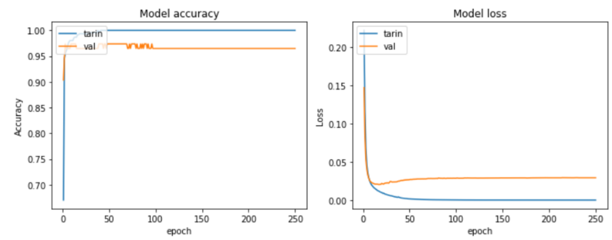

Fig. 14. ANN - MLP (Optimizer: Adam, Loss Function: Mean Squared Error)

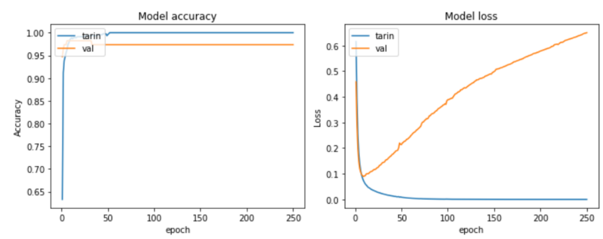

Fig. 15. ANN - MLP (Optimizer: Adam, Loss Function: Binary Cross-Entropy)

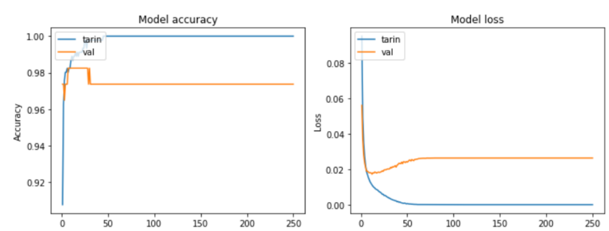

Fig. 16. ANN - MLP (Optimizer: rmsprop, Loss Function: Mean Squared Error)

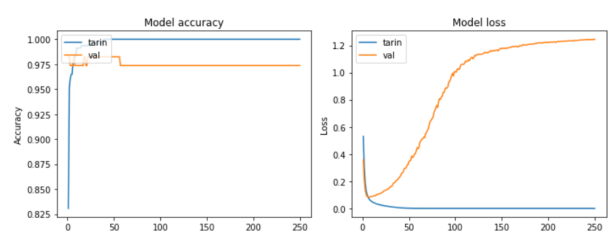

Fig. 17. ANN - MLP (Optimizer: rmsprop, Loss Function: Binary Cross-Entropy)

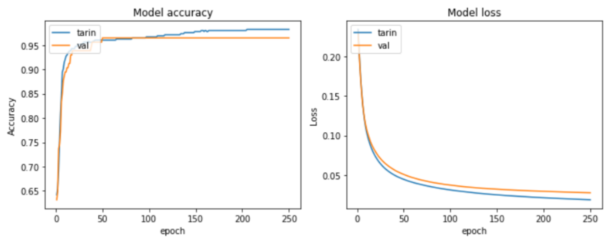

Fig. 18. ANN - MLP (Optimizer: sgd, Loss Function: Mean Squared Error)

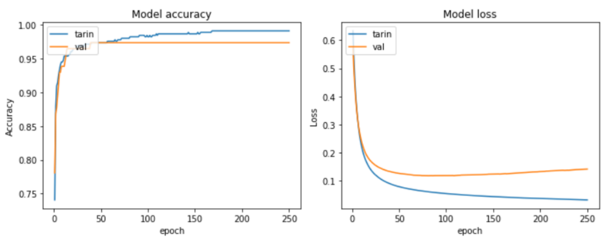

Fig. 19. ANN - MLP (Optimizer: sgd, Loss Function: Binary Cross-Entropy)

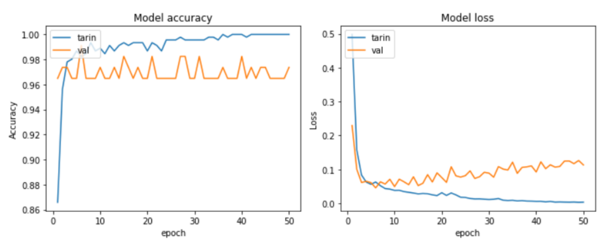

Fig. 20. ANN - MLP w/o Drop-Out and Batch-Normalization

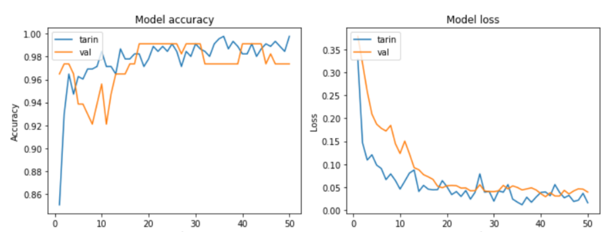

Fig. 21. ANN - MLP w/ Drop-Out and Batch-Normalization

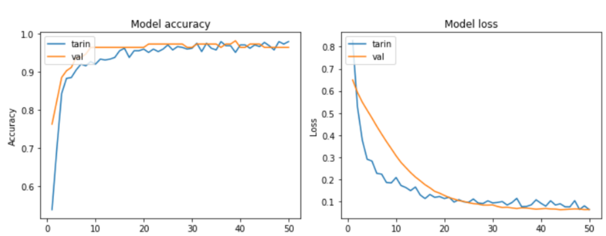

Fig. 22. ANN - MLP w/ Drop-Out and Batch-Normalization (slow learning rate)

Experimental results show that model with Convolution layers, having Drop-Out and Batch-Normalization, with slow learning rate is best choice since it provide highest accuracy on validation data.

Moreover we can see from experimental result that ANN with convolution layers require less number of training epochs, we are getting 99% of validation accuracy with only 50 epochs, while ANN-SLP and ANN-MLP require 200 epochs to reach more than 97% validation accuracy.

REFERENCES

[1] *https://en.wikipedia.org/wiki/Breast*$_c$*ancer*

[2] *https://en.wikipedia.org/wiki/Artificial*$_n$*eural*$_n$*etwork*

[3] UCI Machine Learning Repository *https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin +%28Diagnostic%29*

[4] A. K. Jain, Jianchang Mao and K. M. Mohiuddin, ”Artificial neural networks: a tutorial,” in Computer, vol. 29, no. 3, pp. 31-44, March 1996.

[5] *https://keras.io/*

[6] W.S. McCulloch and W. Pitts, “A Logical Calculus of Ideas Immanent in Nervous Activity,” Bull. Mathematical Biophysics, Vol. 5, 1943, pp 115-133.

[7] The Machine Learning Dictionary”. www.cse.unsw.edu.au. Archived from the original on 26 August 2018. Retrieved 4 November 2009.

[8] Zell, Andreas (2003). ”chapter 5.2”. Simulation neuronaler Netze [Simulation of Neural Networks] (in German) (1st ed.)

[9] DAWSON, CHRISTIAN W (1998). ”An artificial neural network approach to rainfall-runoff modelling”. Hydrological Sciences Journal. 43 (1): 47–66. doi:10.1080/02626669809492102.

[10] H.D.Cheng, Juan Shan, Wen Ju, Yanhui Guo ,Ling Zhang. ”Automated breast cancer detection and classification using ultrasound images: A survey”. Pattern Recognition Volume 43, Issue 1, Pages 299-317, January 2010.

[11] Nilashi, M., Ibrahim, O., Ahmadi, H., Shahmoradi, L. (2017). A knowledge-based system for breast cancer classification using fuzzy logic method. Telematics and Informatics, 34(4), 133–144. https://doi.org/10.1016/j.tele.2017.01.007

[12] F. A. Spanhol, L. S. Oliveira, C. Petitjean and L. Heutte, ”Breast cancer histopathological image classification using Convolutional Neural Networks,” 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, 2016, pp. 2560-2567.

[13] M. Saritas and A. Yasar, “Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification”, IJISAE, vol. 7, no. 2, pp. 88-91, Jun. 2019.

[14] Lam, S. W., Jimenez, C. R., Boven, E. (2014). Breast cancer classification by proteomic technologies: Current state of knowledge. Cancer Treatment Reviews, 40(1), 129–138. https://doi.org/10.1016/j.ctrv.2013.06.006

[15] Abdel-Zaher, A. M., Eldeib, A. M. (2016). Breast cancer classification using deep belief networks. Expert Systems with Applications, 46, 139–144. https://doi.org/10.1016/j.eswa.2015.10.015