Enhancing Multi-Layer Perceptron For Breast Cancer Prediction

3 authors:

Bassam Al-Shargabi
Middle East University
41 PUBLICATIONS   **194** CITATIONS

SEE PROFILE

Feda Sh
Middle East University
2 PUBLICATIONS   **1** CITATION

SEE PROFILE

Rami Suleiman Alkhawaldeh
University of Jordan
30 PUBLICATIONS   **123** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project   A Cultural E-Government Readiness Model View project

Project   Enhanced Chaotic Maps for Cryptographic Applications View project

ENHANCING MULTI-LAYER PERCEPTRON FOR BREAST CANCER PREDICTION

Bassam Al-Shargabi[1*], Fida'a Al-Shami[2] and Rami S. Alkhawaldeh[3]

[1]Computer Information Dept, Middle East University, Amman, Jordan
[2]Computer Science Dept, Middle East University, Amman, Jordan
[3]Computer Information Systems Dept, The University of Jordan, Aqaba, Jordan
[1*]bshargabi@meu.edu.jo, [2]fida_alshami@hotmail.com, [3]r.alkhawaldeh@ju.edu.jo

Abstract—**Breast cancer (BC) is a standout disease of the most well-known cancers among women around the world. The analysis and prediction of BC leads to early manage the disease and protect the patients from further medical complications. In the light of its noticeable focal points in basic highlights identification from complex BC datasets, Machine Learning (ML) is generally perceived as the technique of decision in BC design order and gauge displaying. Because of the high performance of the Multi-layer Perceptron (MLP) algorithm as one of the ML techniques, we conducted experiments in order to enhance the accuracy rate of MLP by tuning its hyper-parameters along with studying the effect of feature selection methods and feature reduction of MLP. As feature selection results indicated that an increase in the number of input parameters tends to reduce the error associated with the estimator model. The tuned MLP proposed in this paper, based MLP best fit hyper-parameters along with feature selection is applied for breast cancer classification using Wisconsin Diagnostic Breast Cancer (WDBC) dataset. As the tuned MLP experimental result shows an accuracy reached 97.70% as it outperforms the basic MLP.**

1. INTRODUCTION

According to the American Cancer Society (ACS), worldwide studies reported 627,000 deaths from breast cancer in 2018 and this number is expected to increase impacting 2.1 million women in 2019. Breast cancer has turned into a typical disease around the globe caused 22.9% of a wide range of cancers in women and the main source of cancer death [1]. Furthermore, about 10% of women have breast cancer in western nations, millions of women experience the ill effects of this weakening dangerous disease [2].

The development of massive breast cancer screening leads to earlier diagnosis and rapid management with a significant improvement in survival rate. The treatment and analysis of medical images is a rapidly expanding area. The problem of automatically searching for information contained in medical images is needed, due to the diversity of medical imaging devices and the difficulty of interpretation of such images [3]. In order to process this large volume of information, doctors are currently turning to the use of systems to assist in the analysis and interpretation of these images. This analysis aims to

facilitate the diagnosis made by the practitioner and to make it as accurate and reliable. This can be achieved by Machine learning (ML) techniques, although breast cancer analysis remains real public health and a very sensitive problem. Therefore, the necessity for accurate ML models is required.

Machine learning (ML) is one of Artificial Intelligence (AI) fields that imitates the human brain in processing data and creating patterns for using in decision making[4]. ML techniques endeavor to develop supervised prediction models for steadily classifying future occasions with historical information. These techniques have often shown up in different practice regions and particularly health-care and biology[5]. Classification and prediction issues have an essential role in medical decision making[6]. In this manner, because of illnesses diagnosis significance to humankind, a few examinations have been led on modeling procedures for their classification [6, 7].

In this study, we used MLP technique as a prediction model for the breast cancer tumors , as we conducted a set of experiments on a variant parameter of the MLP model to enhance the prediction accuracy of the generated model. This research aims to study, as an enhancing, one of machine learning techniques in order to classify or predict breast cancer. With a plethora of techniques, each of which owns a process for building the predictive model. In addition to the importance of detecting the tumour in the early stages, the necessity to find sub-optimal models is required. This research presents the results of the MLP method that would be baselines for researchers to focus more on improving such techniques for the future of breast cancer diseases.

The paper is organized as follows: Section 2 discusses background information about MLP technique and related works in improving its parameters on some applications. Section 3 proposes the methodology in tuning the MLP hyper-parameters. Section 4, the experimental results and discussions were presented along interpretations to the constructed models. Section 5, concludes the final results and exhibits the future work.

2. BACKGROUND AND RELATED WORK

In this section, we present and discussed the main concepts of MLP, feature selection, and grid search, as they are the backbone of this paper.

2.1. MULTILAYER PERCEPTRON

By training on a dataset, the supervised learning algorithm MLP, comes under the functions classifier category that learns a function $f(.): R^m \rightarrow R^o$

where m, is the number of dimensions for input and o is the number of dimensions for output. MLP can be one or more a non-linear layer, which is different from logistic regression, when a set of features $X = X^1, X^2, \ldots \ldots, X^m$ is given and a target **y**, MLP can learn non-linear function approximator for either classification or regression. Figure 1 shows one hidden layer between the input and the output layer, with scalar output.
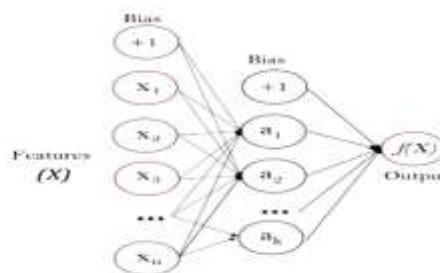

Fig. 1   One hidden layer MLP [8]

Advantages: To use for regression and mapping, the MLP algorithm is a generally very good algorithm. It tends to be utilized to map an N-dimensional input signal to an M-dimensional output signal, this mapping can likewise be non-linear.

Disadvantages: The principal restriction of the MLP algorithm is that the number of Hidden Neurons must be set by the user, setting this value too low may result in the MLP model underfitting while at the same time setting this value too high may result in the MLP model overfitting. Another restriction of the MLP algorithm is that due to how it is trained, it can't ensure that the minima it stops at amid training are the global minima. The MLP algorithm can stall out in local minima [9].

2.2. FEATURE SELECTION

Because of a lot of data streaming over the network real-time machine learning is practically generally complex. Feature selection can reduce calculation time and model complexity. Research on feature selection began in the early 60s. Essentially, feature selection is a strategy of choosing a subset of pertinent/vital features by evacuating the most irrelevant and redundant features from the data for structure a viable and proficient learning model.[10] As shown in Figure 2.
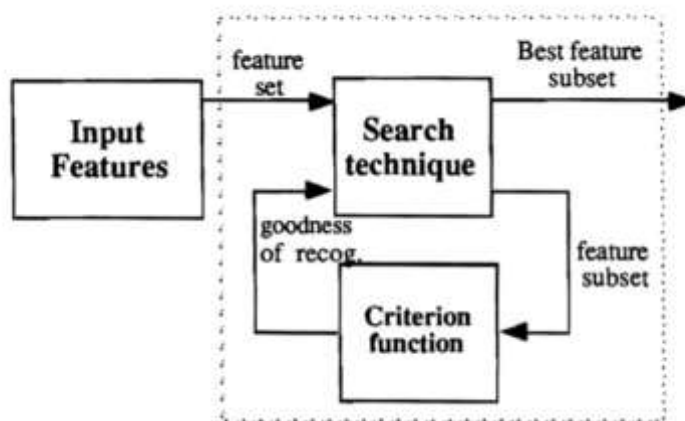

Fig. 2   Block diagram of the adaptive feature selection process [11]

The procedure of Feature Selection forms, as shown in Figure 3, includes four fundamental strides in a run of the mill feature selection strategy that appeared in Figure 2. First is age strategy to produce the following competitor subset; the second one is an assessment capacity to assess the subset and the third one is a halting rule to choose when to stop, and an approval technique to check whether the subset is valid [10].
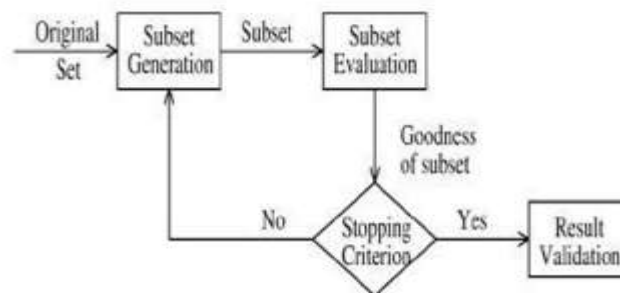

Fig. 3   Feature selection process [10]

2.3. GRID SEARCH

Grid search is the way toward performing hyperparameter tuning to decide the ideal qualities for a given model. This is critical as the performance of the whole model depends on the hyperparameter esteems determined [12]. It works iteratively, for some of the parameters related to the model you enter some great likely qualities and the Grid search emphasizes through every one of them and looks at the outcome for each value and afterward gives you the best parameters which are appropriate for your model [13].

2.4. EVALUATION METRICS

Classification Accuracy is referred to the capacity of the algorithm to anticipate the correct class name for instances of obscure class labels (testing set), as shown Equation 1. The accuracy measure is utilized for evaluating and comparing the underlying basic MLP classifications. [14]

$$classification\ Accuracy = \frac{Number\ of\ correctly\ classified\ samples}{Total\ number\ of\ test\ sampmles} \times 100$$

(1)

2.5. RELATED WORK

This section illustrates several related works to the tuning of MLP and its hyperparameter or other methodologies to improve machine learning.

A hybrid model composed of MLP with either a particle swarm optimization (PSO) algorithm or a cuckoo optimization algorithm (COA) was proposed in [15]. Their approach based on plus-l-take-r for feature selection as it increases the number of parameters used for the training model with the aim of minimizing the error rate of the model. As they concluded to reduce the number of parameters to only five parameters for the prediction model. They conducted experiments to compare MLP-COA and MLPPSO hybrid models, the results shows that MLPPSO outperforms MLP-COA in term of accuracy.

An ensemble based feature selection utilizing random trees and wrapper technique was proposed in [16]. The proposed ensemble learning classification technique infers a subset utilizing the wrapper strategy, bagging, and random trees. The proposed strategy evacuates the unimportant features and selects the ideal features for classification through probability weighting criteria[18,19]. The proposed features selection technique is evaluated utilizing Support Vector Machines (SVM), Random Forest (RF), and Naïve Bayes (NB) evaluators and the exhibitions are thought about against the Feature Selection Naïve Bayes (FSNBb), Feature Selection Support Vector Machines(FSSVMb), Genetic Algorithm and Support Vector Machine (GASVMb), Genetic Algorithm and Naïve Bayes (GANBb), and Genetic Algorithm and Random Forest (GARFb) strategies. The proposed strategy accomplishes mean classification accuracy of 92% and beats the other ensemble techniques.

A hybrid machine learning approach named Multi-layer Perceptron Neural Network (MLP)and Biogeography-based Optimization (BBO (MLP-BBO) for the prediction of the coefficient of consolidation of soft soil was presented in [17]. This technique depends on Multi-layer Perceptron Neural Network (MLP)and Biogeography-based Optimization (BBO). For this, an all-out number of 164 soil tests. In the displaying, input parameters utilized include clay (%), moisture content (%), liquid limit (%), plastic limit (%), plasticity index (%), and liquidity index (%). To approve the execution of the models, standard techniques, for example, Root-Mean-Square Error (RMSE), Mean Absolute Error (MAE), and relationship coefficient (r) were utilized. For the examination of the execution of the models, mainstream machine learning strategies were utilized to be specific Back propagation Multi-layer Perceptron Neural Networks (Bp-MLP Neural

Nets), Radial Basis Functions Neural Networks (RBF-Neural Nets), Gaussian Process (GP), M5 Tree, and Support Vector Regression (SVR). Consequences of the model examination demonstrate that the proposed strategy MLP-BBO has the most noteworthy prescient capacity ((RMSE=0.397, MAE=0.302 and r=0.827), followed by the SVR (RMSE=0.403, MAE=0.299 and r=0.819), the Bp-MLP Neural Nets (RMSE=0.478, MAE=0.398 and r=0.805), the RBF-Neural Nets (RMSE=0.412, MAE=0.281 and r=0.804), the GP (RMSE=0.440, MAE=0.301 and r=0.777), and the M5 Tree (RMSE=0.494, MAE=0.322 and r=0.728), in the prediction of Cv, separately. Along these lines, it very well may be inferred that however execution of every single connected model is good in the prediction of the coefficient of consolidation of soil yet execution of the MLP-BBO is the best.

A knowledge-based system for predicting breast cancer is proposed in [20]. the system uses Expectation Maximization (EM) and Principal Component Analysis (PCA), as they were individually exploited for clustering and addressing the multicollinearity problem within the datasets. Moreover, they used Regression Trees (CART) to consequently create set fuzzy rules from the dataset that will help predict the presence of Brest cancer of mammogram images. They evaluated the knowledge-based system for two real-world datasets, WDBC and Mammographic mass same as related to other studies [21, 22].

3. PROPOSED TUNED MLP MODEL

The classification is a core process for a large number of breast cancer prediction, which requires features that are extracted from dataset. [18] The Tuned MLP extensions are generally divided into two categories, one investigated the reduction of the extracted feature size, while the other investigated the enhancement of the classification power. This section introduces a tuned MLP for local feature extraction by extending the MLP with the aim of reducing the size of the extracted feature and enhance the classification power discriminating against different ML classification algorithms.
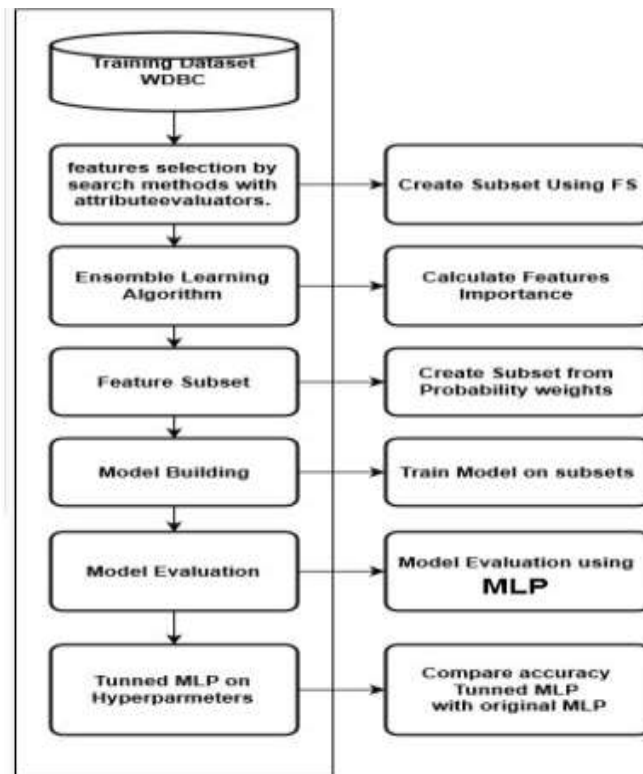

Fig. 4   The Tuned MLP

The proposed model used consists of the following steps:

A. Data Selection and Preparation:

The dataset is selected from WDBC. This dataset includes 569 instances and 32 attributes.

B. Preprocessing Data

In this step the data set is loaded then a supervised filter which is attribute-based is applied on the data set which removes unnecessary attributes and therefore 1 attribute was selected.

C. Classification Iteration 1

In this step, we determine to classify and select cross-validation 10 and select the MLP algorithm.

D. Ensemble feature selection

In step, we define and select attributes and choose a search method and attribute evaluators such as search method;(attribute evaluators): (Ranker; InfoGainAttributeEval), (BestFirst; CfsSubsetEval), (Greedy Stepwise; CfsSubsetEval), (BestFirst; WrapperSubsetEval), (Greedy Stepwise; WrapperSubsetEval).

E. Voted of the best relevant features

As in the previous step, the number of features will be reduced, in this step the voting for the best relevant features through the following equation:

$$Y = w_1c_1 + w_2c_2 + w_3c_3 + w_4c_4 + w_5c_5 \qquad (2)$$

F. Classification using the tuned MLP

By trying and experimenting and comparing the parameters of the MLP algorithm to raise its accuracy in the classification of breast cancer.

That Tuned MLP parameters that we used in the proposed model are as follows, N: the number of the hidden layers, activation function for the first N-1 of the hidden layer, the last hidden layer activation function (is sigmoid because the type of class is binary), the optimizer, hyper-parameter (learning rate=0.001, patch size= 5, number of epochs =150 and dropout to avoid overfitting problem =0.1.

4. EXPERIMENTAL RESULT AND DISCUSSION

In this section, the experimental result conducted for the Tuned MLP through setting the hyper-parameters of basic MLP, and compare the result with tuned MLP along with other ML such as KNN, RF.

4.1. PARAMETER SETTINGS AND DATASET

Before conducting the experiments, a set of parameters for the MLP and dataset were set up as in Table I.

Table I. Parameter Settings of MLP classification algorithms

Parameter	MLP
Input layer	31
Output layer	1
Type of class	**Binary { M = malignant, B = benign }**
dataset	WDBC

The tuned MLP is mainly about the best fit for MLP to improve the accuracy, where several parameters that affect the MLP results. That parameters are as follows, N: the number of the hidden layers, activation function for the first N-1 of the hidden layer, the last hidden layer activation function (is sigmoid because the type of class is binary), the optimizer, hyper-parameter (learning rate=0.001, patch size= 5, number of epochs =150 and dropout to avoid overfitting problem =0.1. The values of parameters for the conducted experiments for basic MLP and its accuracy without ensemble feature selection is shown in Table II.

Table II. Tuning MLP Parameters

seed	activation1	activation2	hidden layer	optimizer	accuracy
2	Relu	sigmoid	1	rmsprop	**97.35%**
2	Relu	sigmoid	1	adem	**96.13%**
2	Tanh	sigmoid	1	rmsprop	**96.65%**
2	Tanh	sigmoid	1	adem	**96.65%**
2	sigmoid	sigmoid	1	rmsprop	**96.47%**
2	sigmoid	sigmoid	1	adem	**97.00%**
2	Relu	sigmoid	2	rmsprop	**97.53%**
2	Relu	sigmoid	2	adem	**96.83%**
2	Tanh	sigmoid	2	rmsprop	**97.18%**
2	Tanh	sigmoid	2	adem	**97.00%**
2	sigmoid	sigmoid	2	rmsprop	**97.18%**
2	sigmoid	sigmoid	2	adem	**97.00%**
2	relu	sigmoid	4	rmsprop	**96.30%**
2	relu	sigmoid	4	adem	**97.01%**
2	tanh	sigmoid	4	rmsprop	**95.61%**
2	tanh	sigmoid	4	adem	**96.65%**
2	sigmoid	sigmoid	4	rmsprop	**97.52%**
2	sigmoid	sigmoid	4	adem	**96.99%**
2	relu	sigmoid	6	rmsprop	**96.30%**
2	relu	sigmoid	6	adem	**96.65%**
2	tanh	sigmoid	6	rmsprop	**96.83%**
2	tanh	sigmoid	6	adem	**96.30%**
2	sigmoid	sigmoid	6	rmsprop	**97.18%**
2	sigmoid	sigmoid	6	adem	**97.70%**

4.2. RESULT AND DISCUSSIONS

At the beginning the first experiments for the basic MLP accuracy result (96.66%) as shown in Table III, then started the tuning process, by implementing extra hyper-parameters for the proposed tuned MLP, to find the best fit that eventually will produce higher accuracy for the prediction of Brest cancer.

Table III. Accuracy results of MLP on the WDBC dataset before the selected feature

Scheme	Accuracy						
	Correctly Classified Instances	Incorrectly Classified Instances	Kappa statistic	Mean absolute error	Root mean squared error	Relative absolute error	Root relative squared error
MLP	96.66%	3.34%	0.93	0.04	0.17	7.82%	34.98%

The second experiment, for the tuned MLP through reduced the feature selection from 31 attributes to only 4 attributes that were most influential by identifying the tumour type. Search methods were implemented with different attribute evaluators, for example, Ranker with InfoGainAttributeEval, BestFirst with CfsSubsetEval, Greedy Stepwise with CfsSubsetEval, BestFirst with WrapperSubsetEval, and Greedy StepwiseWrapperSubsetEval, Then , we voted the selected feature, as the number of feature was reduced from 31 to 4 feature results indicated four features (texture_mean, concave points_mean, Radius_worst, Smoothness_worst) of WDBC dataset that were dominant for determining of presence of Brest cancer.

Table V. Accuracy results of basic MLP algorithm on the WDBC dataset after selected attributes

Scheme	Accuracy						
	Correctly Classified Instances	Incorrectly Classified Instances	Kappa statistic	Mean absolute error	Root mean squared error	Relative absolute error	Root relative squared error
MLP	97.19%	2.81%	0.94	0.04	0.15	7.80%	31.03%

The third experiment were conducted by exploiting the use of Grid Search method, which is used to find the optimal hyperparameters (as shown Table II)of a model which results in the most 'accurate' predictions, to find the best value of the parameters for the optimizers in terms of epochs, batch-size and seed. The result of experiments is evaluated based on which of the Grid Search method produces the best result in terms of accuracy as shown in Figure 5.

As noted in Figure 5, the best parameters that produce accuracy (97.70%), which is better as compared to the basic MLP (96.66%) as discussed early in this section. The best fit for the Tuned MLP as proposed in this paper is as follows: as the activation function for the first N-1 Hidden layers is sigmoid, the number of hidden layers is six, the optimizer is Adem.
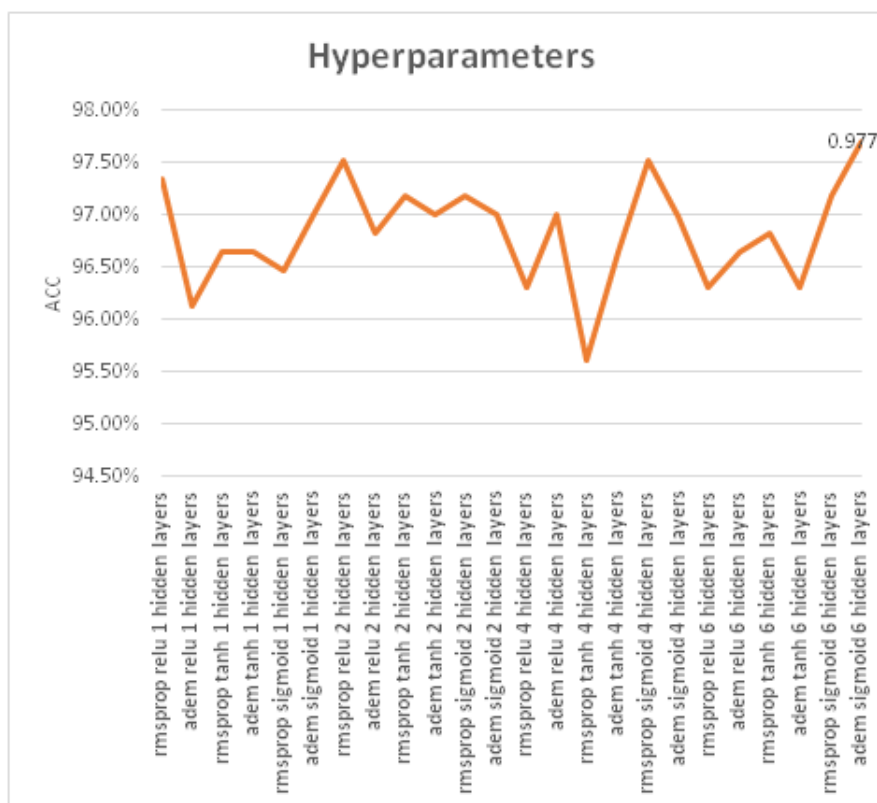
Fig. 5   Experiments for Tuned MLP Parameters

5. CONCLUSION

In this paper, a tuned MLP   for breast cancer prediction is proposed, which is aimed at improving the accuracy of MLP compared to the basic MLP algorithm. The Tuned MLP is based on reducing the size of the extracted feature, as we used search methods were used with different attribute evaluators, and then voted of the best relevant features where it was reduced features from 31 attributes to only 4 attributes of WDBC dataset.

Moreover. The proposed Tuned MLP proven that the use of Grid Search method to find the optimal hyperparameters of a model which results in the most 'accurate' predictions as compared to the basic MLP. The tuned MLP best fit for hyperparameters are as follows: as the activation function for the first N-1 Hidden layers is sigmoid, the number of hidden layers is six, the optimizer is Adem as the experimental results proven over the WDBC dataset showed that the accuracy enhancement of tuned MLP compared to the original MLP is around 1.04%.

ACKNOWLEDGMENTS

REFERENCES

[1]   Duijm, L. E. M., Groenewoud, J. H., Jansen, F. H., Fracheboud, J., van Beek, M., & de Koning, H. J. (2004), "Mammography screening in the Netherlands: delay in the diagnosis of breast cancer after breast cancer screening", British journal of cancer, 91(10), 1795.

[2]   Pendharkar, P. C., Rodger, J. A., Yaverbaum, G. J., Herman, N., & Benner, M. (1999), "Association, statistical, mathematical and neural approaches for mining breast cancer patterns", Expert Systems with Applications, 17(3), 223-232.

[3]   Zheng, Y., Yu, J., Kambhamettu, C., Englander, S., Schnall, M. D., & Shen, D. (2007, October), "De-enhancing the dynamic contrast-enhanced breast MRI for robust registration", In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 933-941), Springer, Berlin, Heidelberg.

[4]   Wiki, A. I., "Artificial Intelligence (AI) vs. Machine Learning vs. Deep Learning", skymind. [Online]. Available: https://skymind.ai/wiki/ai-vs-machine-learning-vs-deep-learning.

[5]   Deo, R. C., "Machine Learning in Medicine", Circulation, vol. 132, no. 20, pp. 1920–1930, Nov. 2015.

[6]   Karabatak, H., "A new classifier for breast cancer detection based on Naïve Bayesian", Meas. J. Int. Meas. Confed., vol. 72, pp. 32-36, 2015.

[7]   Zare-Zardini, H, "Nanotechnology and Pediatric Cancer: Prevention, Diagnosis and Treatment", Iran. J. Pediatr. Hematol. Oncol., vol. 5, no. 4, pp. 233–248, 2015.

[8]   scikit-learn.org, "Neural network models (supervised)", [Online]. Available: https://scikit-learn.org/stable/modules/neural_networks_supervised.html#complexity. [Accessed: 25-Mar-2019].

[9]   [NickGillian, 'GRT: MLP', 2014.

[10] Aggarwal, M., "Performance analysis of different feature selection methods in intrusion detection", Int. J. Sci. Technol. Res., vol. 2, no. 6, pp. 225–231, 2013.

[11] Haleh, V., & Imam Ibrahim, F., "Feature Selection Methods : Genetic Algorithms vs. Greedy-like Search", Proc. Int. Conf. Fuzzy Intell. Control Syst., vol. 1, no. Vafaie 93, pp. 1-10, 1997.

[12] Hewa, K., "An introduction to Grid Search", Data Driven Investor, 2019. [Online]. Available: https://medium.com/datadriveninvestor/an-introduction-to-grid-search-ff57adcc0998. [Accessed: 01-Jan-2019].

[13] Singh, A. V., "What is a grid search, and why do we use it in machine learning?", 2018. [Online]. Available: https://www.quora.com/What-is-a-grid-search-and-why-do-we-use-it-in-machine-learning. [Accessed: 01-Apr-2019].

[14] Suleiman, A. A., "Image Classification Based on Enhancement of Local Binary Pattern", MEU, 2018.

[15] Anemangely, M., Ramezanzadeh, A., Tokhmechi, B., Molaghab, A., & Mohammadian, A. (2018), "Drilling rate prediction from petrophysical logs and mud logging data using an optimized multilayer perceptron neural network", Journal of Geophysics and Engineering, 15(4), 1146-1159.

[16] Mohan, C., & Nagarajan, S. (2019), "An improved tree model based on ensemble feature selection for classification", Turkish Journal of Electrical Engineering & Computer Sciences, 27(2), 1290-1307.

[17] Pham, B. T., Nguyen, M. D., Bui, K. T. T., Prakash, I., Chapi, K., & Bui, D. T. (2019), "A novel artificial intelligence approach based on Multi-layer Perceptron Neural Network and Biogeography-based Optimization for predicting coefficient of consolidation of soil", Catena, 173, 302-311.

[18] Zheng, B., Yoon, S. W., & Lam, S. S. (2014), "Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms", Expert Systems with Applications, 41(4), 1476-1482.

[19] Sabri, A.T., Mohammads, A.H., Al-Shargabi, B. and Hamdeh, M.A., 2010, "Developing new continuous learning approach for spam detection using artificial neural network (CLA_ANN)", European Journal of Scientific Research, 42(3), pp. 525-535.

[20] 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.

[21] Aljawarneh, S., Al-shargabi, B., Rashaideh H., (2013), "Gene classification: A review", In Proceedings of IEEE ICIT'2013 May 8.

[22] Al-Shargabi, B., Amro, I, Kanaan. G. (2009), "Exploit Genetic Algorithm to Enhance Arabic Information Retrieval", In 3rd International Conference on Arabic Language Processing (CITALA'09), Rabat, Morocco, May 2009.