APPLICATION OF MACHINE LEARNING MODELS AND PARAMETERS FOR PREDICTION OF BREAST CANCER

A.A. Malik[1], K. Amjad[2], A. Saud[3], S. Aslam[4], A. Ali[5]
[1,3,4,5] Faculty of Computer Science Lahore Garrison University
[2] Department of Computer Science, University of Management and Technology, Lahore
Corresponding Author's Email: dr_aftab_malik@yahoo.com

ABSTRACT: Breast cancer is the only type of cancer which is affecting the woman on a large scale because the tumors which are developed in the early stage of cancer or the origin of cancer these tumor cells or lumps are divided into two different kinds and both are dangerous in their own away as one of the lump is cancerous and other is not but both of them are one of the main reason behind the increasing numbers of death among the women. The purpose of the proposed research is control the rising tally of death day by day all over the world because there no efficient way of predicting or diagnosing the tumor of breast cancer whether the lumps developed in the patients are cancerous or not so, the objective of the proposed research is to overcome this major problem by using machine learning and deep learning as machine learning (ML) can provide certain tools which can be used in addition with hyper parametric for diagnosing the tumor in an efficient way that is the main reason behind the use of six supervised Machine Learning(ML) algorithm involving . The KNN, SVM, DT, and DP methods like Adam Gradient learning were utilized due to their adaptive gradient algorithm and root mean square propagation benefits; as in each model is has been proved that it shows more accuracy individually in the models or on comparing them with each other as the accuracy shown by these method was 98.50%. as described in result and discussion section.

Key words: Breast cancer, Tumor cell, Machine learning, Deep learning, Algorithms, Mammogram Analysis.

(Received 11.09.2023 Accepted 12.11.2023)

Abbreviations used:

a. ANN: Artificial Neural Networks
b. SNN: Simulated Neural Networks
c. KNN: K-Nearest Neighborhood
d. SVM: Support Vector Machine
e. DT : Decision Tree
f. DP : Deep Learning

INTRODUCTION

Breast cancer is a prevalent disease, with early detection through mammography and fine needle aspiration, and treatment varying based on stage, with surgery being the primary treatment, because the size of the tumor is less than two inches and it is better to remove the tumor. Similar to stage II of breast cancer the size of the tumor slowly grows largerthan in the first stage and to stop the tumor from growing surgery and chemotherapy are the most preferred treatments. Moreover, at stage III the size of the tumor is mostly five centimeters and chemotherapy with or without targeted therapy is done first followed by surgery and radiotherapy. Furthermore, hormone therapy is also considered if the cancer cells have hormone receptors. The serious stage of breast cancer is termed as stage IV , when cancer sells rapidly grow and spread to other parts of the body and at this stage treatment method such as hormone therapy, chemotherapy, and immunotherapy drugs are used to treat the convicted patients. Furthermore, after early identification of cancer using the proposed method the convicted patient can be treated using these methods according to the stage of the breast cancer; for identification of cancer using the proposed method the convicted patient can be treated using these methods according to the stage of the breast cancer. All the above methods are conducted by specialist surgeons.

Breast Cancer is a disease or it can be called a heterogeneous disease that starts when the tissue in the breasts starts growing and the growth of these cells or tissues is abnormal as they grow out of the control. This abnormal growth of cells forms a sheet known as a tumor that spreads slowly in the whole body of the patient and it is also one of the most important symptoms while diagnosing breast cancer in patients. Both men and women can suffer from this disease but most women are being diagnosed with breast cancer according to the

collected data about more than two hundred eleven thousand cases of breast cancer were reported in 2005.According to the research breast cancer is mostly found in a woman above 40 and about 85% woman's of above 50 are affected by breast cancer [1]. Breast cancer slowly evolves in the body and then it slowly spread and infects the whole body this reason why most people do not notice it and most people divorce it in their regular screening. Breast cancer can also be discovered through proper physical examination, imaging or the most affected method is mammography because it helps in identifying whether the lumps are cancerous or not. All the lumps are not cancerous and about 80 % of a woman were diagnosed with this disease through mammograms thus there are many other methods to diagnose breast cancer [2].

which is used for diagnosing breast cancer Fine Needle

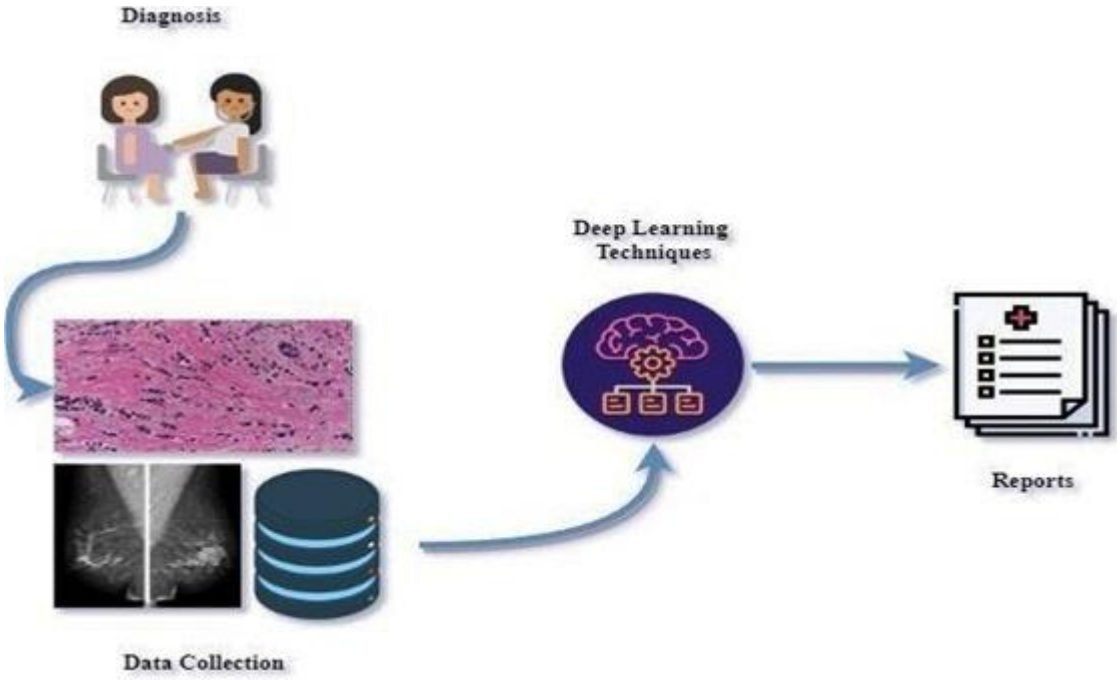

Figure 1 Breast Cancer System Workflow

Aspiration is a biopsy in which a small needle of 25 gauge is passed through the skin of the patient to collect a sample of the solid mass which is then examined in the pathology laboratory. The method of Fine Needle Aspiration is used when an unwanted lump is found in a patient while going through different tests like x-rays, ultrasound, or mammography the fine needle aspiration is used because it is less painful and a quicker method for sampling the tissue as compared to other methods [3-5].

Fine-Needle Aspiration is an effective method for identifying the key feature of lumps such as the radius, size, and area covered by the lumps. These features can be identified through these methods and can be dived into two different types one of them is Benign which is a non-cancerous lump and the other is Malignant which is a cancerous lump. The benign is less dangerous as compared to the malignant cells because these types of cells do not spread and they do not attack the tissue which is next to them they do not spread in the body but in case of any connection with blood vessels or nerves. These cells can be very dangerous on the other hand the malignant cell grows slowly and spread in the body as they keep on surrounding the tissues which are next to them [6]. Machine learning (ML) provides the idlest framework that gives the required support for finding or predicting the levels of vulnerability to which a malignant sickness can grow. The results, however, depend upon the provided dataset and the information machine learning (ML) uses statistical techniques to determine the growth of breast cancer the most common algorithm is k-nearest neighbor which is also a learning technique utilized for k-nearest forecasting the medical trends. Thus, by using the method of Fine needle aspiration about 80% of the patients with this disease were detected. The novelty of the proposed research is that by using the Machine learning (ML) method or techniques for the prediction of breast cancer about 99% of patients with this disease can be detected. Machine learning (ML) methods can also be used for the diagnosis of both cells (Malignant and benign). In the proposed research Wisconsin Breast cancer dataset can access the bosom malignant growth dataset and is generally examined [7-8].

Related Work: The researcher used several different algorithms of machine learning (ML) to categorize the different types of tumors during this process a small part of the neural networks of the short-term memory which are also known as a gated recurrent unit (GRU) in this process the support vector layer was used instead of the soft-max layer as the combination of these two methods have the highest accuracy that is shown in Table I.

In another research, the researcher used the method of Waikato Environment for knowledge analysis (WEKA) in combination with the SVM, and the best results were obtained from it. A computer-aided designed system of liner analysis was used with SVM and it has an accuracy of 99% [9-10].

A researcher named Kumari used a new technique of Machine learning to identify different unknown patterns of the database data set of WBCD, for predicting breast cancer using a different method of machine learning. The Association rules with the neural networks were applied by the researcher named Karabatak; Cross-validation was utilized by him to train a specific model, enhancing its precision similar to the Naïve Bayes classifier application process. From table 1 it can be determined that entirely machine learning models the way the Naïve Bayes classifier was applied.

In [26], the authors from the past conducted many different experiments like recursive least square that provides maximum value of Correct Classification rate to (CCR) from 0.8 and 0.9 for more than one predictor. For these methods the results are obtained in optimum time and comparatively with less computational work as well as less description length. From the table I it can be determined that entirely machine learning models can be used for the diagnosis of the cells and all of them provide almost the same results and accuracy therefore it is difficult to go for an individual model of technique which is also the main objective behind the parametric study of these different models in addition to the machine learning (ML) and Deeping learning (DP) which are also applied for the categorizing these models the uniqueness of each model depends upon it is being implemented.

Table 1 Comparison of Proposed system with Current Systems

SR.	Machine Learning Model	Accuracy Of Models	Published Year
01	GRU, SVM. Liner Regression, KNN, Soft max regression	99.18%	2018
02	SVM, Naive Bayesian, KNN	98.08%	2016
03	LDA Infused With SVM	99.03%	2019
04	SVM and Relevance Vector Machines	99.42%	2018
05	BST, ANN, SVM	97.60%	2015
06	Decision Tree	71%	2017
07	Association Rules and NN	95.12%	2004
08	Naive Bayesian	90.55%	2016
09	Ensemble Method	89.8%	2017
10	RVM, SVM, NN	98.50%	2013
11	RBFNN	99.70%	2018

From the table I it can be determined that entirely machine learning models can be used for the diagnosis of the cells and all of them provide almost the same results and accuracy therefore it is difficult to go for an individual model of technique which is also the main objective behind the parametric study of these different models in addition to the machine learning (ML) and Deeping learning (DP) which are also applied for the categorizing these models the uniqueness of each model depends upon it is being implemented.

METHODOLOGY

A. Dataset: The proposed research categorizes the dataset into malignant and benign tumors. For the prediction of breast cancer, another dataset known as "Cancer-Wisconsin dataset" is used from UCI machine learning (ML) repository containing 570 instances and 32 attributes was used for the computation of the features. The features of breast mass were computed using a digitalized image from a fine needle aspirate (FNA). The data set used in the research is multivariate with real-time attributes. The features are obtained from every nucleus and provide a separate nucleus in the image as there is no null value of the features. Area, parameter, Radius, texture compactness, concave point, and symmetry. Concavity is some of the key features. The features are used in the prediction of breast cancer.

B. System Architecture: The architectural diagram shows the relationship between different entities of the system. To detect breast cancer in a patient in the early stages, mammogram images are used. The convicted patient has to undergo mammography after which the mammogram analysis is performed to determine the type of tumor **(Benign and Malignant).** Machine learning techniques are used to classify if the tumor is cancerous

or not or if there is uncertainty in the results. If there are any signs of abnormal growth of tissues (lumps) then the system will further detect the most suspicious regions in the mammograms and in case of uncertainty pathology analysis will be performed to attain more information about the images of pathology reference. A comprehensive report will be generated by the system at the end.

Table 2: Key features for Predication

Features Considered for Breast Cancer Detection
Wisconsin dataset
Area, parameter
Radius, texture.
Compactness, concave point, symmetry

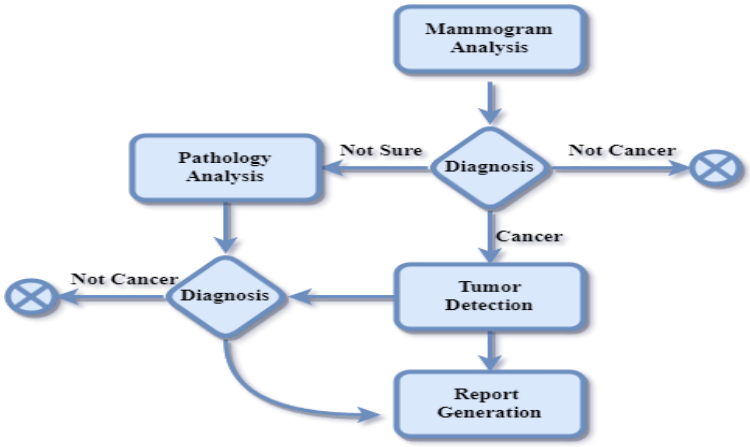

Fig.2 Architectural Diagram

C. Dataset Pre-Processing: The Equation Below was used to standardize the dataset

$$Z = Y - \mu/\sigma \qquad (1)$$

Where Y donates the features which are to be standardized, μ is donating the mean and σ is a deviation standard [11-12].

There are two useful functions of Python we will apply in our scheme, first one is StandardScaler() perform scaling of features as it is a crucial step in modeling algorithms with datasets, derived through various methods. The other function is fit_transform() method, which efficiently fits data into a model and transforms it into a more suitable form, saving time and effort by calling both functions separately.

D. Machine Learning (Ml) Algorithms: In this respective study, the algorithms of machine learning remained available to categorize the two tumor cells which are Malignant and benign and the researcher also studied the parametric aspect of these six models of machine learning (ML) algorithms table 1 above demonstrations the details of machine learning models, with their principal parameters. The machine learning model which has been used in our scheme.

4. Algorithms used

1) Description of SVM): This is algorithm is a supervised learning mechanism and is most popular in working with machine learning, used to solve problems related to "regression" and "classification". In the SVM model the non-linear dataset is converted into linear space which can be further divided into n-dimensional interplanetary areas where a hyperplane of separate classes can be initiate in this research the radically based kernel is used which can be seen in the equation (3) below

$$K\,(X1 - X2) = \exp\,(-\gamma||X1 - X2||2) \qquad (3)$$

2) Decision Tree (DT): Decision trees are used for both classifications and regression problems but like SVM decision trees are mostly used to solve classification problems. Iterative Dichotomiser (ID3) is one of the most popular algorithms which demonstrates the information gained from the attributes to construct the trees. Max-depth, min –samples-leaf, and maximum leaf-node s are the parameter which is being used in the research the equations used in the decision tree are

$$\boldsymbol{Entropy = -pilog2pi} \qquad \textbf{... (4)}$$

$$\boldsymbol{Information\ Gain = Entr(T) - \sum ||ssv||\ Entrpoy(Sv)} \qquad \textbf{(5)}$$

3) Description of KNN

K-Nearest Neighbor (KNN) is a non-parametric model of machine learning in which no assumptions are made it is also known as the **Lazy Learner Algorithm** this is because the K-nearest algorithm stores the dataset when classification is done it performs the required action on it instead of training the set that's why it is considered as the simplest machine learning algorithm based of the supervised learning method. In this process. The process selects the nearest neighbors based on the Euclidean distance calculated between the x and y vectors. The result of the K-Nearest Neighbor(KNN) depends upon the higher k value can be led toward the overlapping in the classes whereas the lower k value will increase the computations [13-14].

$$\text{Euclidean Distance}=\sqrt{\sum_{i=1}^{k}(xi-yi)^2} \quad (6)$$

4) Artificial Neural Networks (Ann and Deep Learning): The ANN and SNN are the subset of the machine learning algorithm (ML) because of which they can also be called the heart of deep learning. The supervised classification problem employs a backpropagation algorithm, contrasting with the Adam Gradient Descent cost function, AdaGrad, and RMSprop.

The following are the parameters that are used in deep learning: The Activation function of ReLU

$$(x) = \max(0, x) \quad (7)$$

The next step is to apply to convert the output into several classes The number of propagated objects through the networks will decide the size of the batch. The system's performance is assessed through binary cross entropy loss. and to increase the predicated probability to show an increase in the actual value [15-16].

5) Random Forest (RF): Random Forest (RF) is a type of machine learning algorithm which is based on the concept of Ensemble

Leaning: According to [17-19] Random Forrest (RF) is a valuable Supervised Learning method which can be equally applicable for regression and classification. The following parameters are used in RF approximates, guesses and defines the number of trees used max_features: total attributes which are to be selected for data randomization, max_depth: for pre-pruning of trees, max_features=sqrt (features) is used for the classification.

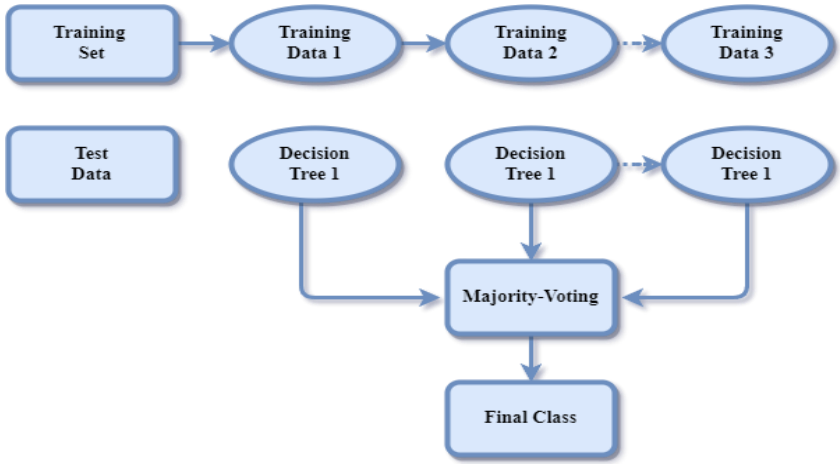

Fig. 3. Working of Random Forest (RF) Algorithm

6) LOGISTIC REGRESSION (LG): Logistics Regression (LG) is a supervised machine learning algorithm (ML) widely recognized as one of the most effective approaches in machine learning. Logistics Regression gives the output in the form of discrete values which are between 1 and 0 or it can be said that the logistics regression (LG) predicts the dependents variables by using a set of independent variables It can be applied using two parameters L1 and L2 as all the feature so only this parameter will be applied to the L2 Regression or

Ridge Regression can be calculated using the equation below

$$L2(C) = w^{*}=\text{argmin}\sum_i \ln[(1\ \exp(-zi))] + \pi * \sum(wj)^2 \ldots (8)$$

Where,

$\sum(wj)^2$ is used for regulation term

π donates the hyper parameter $\sum_i \ln[(1\ \exp(-zi))]$ donates the loss factor;

C is known as the coefficient of regularization

RESULT AND DISCUSSION

A system with 33 GB RAM with an intel processor of inter-core i5. Hex is a modern data platform that enhances Jupyter notebooks by allowing users to connect to any data source, use any Jupyter library, and create interactive data apps therefore, we use 8th Generation python 3.0. The following library was also

used during the experiment Pandas, matplotlib, Tensor flow, and Keras [20], [21].

The limitation in predicting breast cancer is the training of the dataset for the creation of a particular model which can address all the known risk factors and issues and provide the accurate result simultaneously. Current prediction models focus on the analysis of mammographic image risks without other critical elements. These models have the ability to identify high-risk women accurately however the multiple sampling with magnetic resonance imaging (MRI) and ultrasound can affect the patient financially and psychologically. The limitation of magnetic resonance imaging (MRI) for breast cancer is to distinguish between cancerous abnormalities this can be a validity threat as it can result in an unnecessary breast biopsy. All the models were evaluated according to the statistics obtained after measuring the recall, accuracy and precision, recall, and the F1 score obtained using Equations (9 to 12) as given below:

Accuracy= (TruePositiveTrueNegative)/(Tp+Fp+FN+TN) … (9)

Precision=TruePostive/(TP+FP) ... (10)

Recall=Truepostive/(TP+FN) … (11)

F1Score= 2*(Recall*Precision)/(Recall+Precision) …. (12)

Where,

TP=True Positive; TN=True Negative

FP=False Positive

FN=False Negative (False Negative Rate from confusion Matrix)

A. Support Vector Machine (SVM)

Table 3: Result Of Different Parametric

Parameter	Accuracy
C=0	.629
C=1.0	.951
C=100	.972

The 'C' shown in Table III above is donating the regularization Parameter with radical Basis Function and the value of C is being evaluated by giving three different values respectively and the maximum accuracy was obtained when the value of C was equal to 100 which can be seen above in the table.

B. Decision Tree (DT)

Table.4: Result of Decision Tree (DT) Against Different Depths

Depth	Accuracy
Normal Depth=2	.953
Maximum Depth=4	.958

To check the accuracy of the decision tree (DT) different depths were used and the decision tree with a depth of 4 has the highest accuracy [22-23].

C. K-Nearest Neighbor (KNN)

Table 5 Measuring The Accuracy While Varying The Value of 'K'

Neighbors	Accuracy
K=3	.890
K=5	.930
K=6	.958

In the experiment, three different values were used and the highest accuracy found was at k=6 which is demonstrated in table IV shown above.

D. Artificial Neural Networks (ANN and DEEP LEARNING)

Table 6: Different Values of Epochs

DL_AN N	Accuracy Loss
Epoch=1	.5605
	.6928
Epoch=150	.9902
	.0419

In the following experiment, the neural network was trained from epoch=1 to epoch=150 and the objective was to enhance accuracy and minimize loss, but the results of deep learning vary based on the number of epochs, as shown in table 4.

E. Random Forest (RF)

Table 7: Random Forest Behavior Against Different number of trees

n_estimators	Accuracy
10	.890
100	.972
150	.950

From the data obtained from the experiment, it can be seen that an estimator of 100 gives a higher accuracy so a total of 100 decisions were selected.

F. Logistic Regression (LG)

Table 8: Assessment Based On 'C'

Value of the Coefficient C	Accuracy
C=10	.930
C=1	.958
C=100	.965

In the experiment above the algorithm LG has been used with different values of' C and the variation in the value of C can be seen and the most accurate value of C was found at C=100.

In table above different models which were used during the research are compared on the basis of the result obtained from the experiments which is demonstrated in the table 8 above and on the analysis of the obtained result it can be clearly seen that all the models differ from each other on the basis of their accuracy, precision and other factor but among all the algorithms the method of deep learning with the artificial neural network came out as the was accurate method with accuracy of 98.28% which is the highest as compared to other models as mentioned earlier the difference between these model is very small and it can also be seen from table as both SVM and the RF have the same accuracy of 97% on other K-Nearest Neighbors (KNN) and logistics Regression (LG) are the two model which are least accurate. The SVM is a preferred choice for complex datasets, including linear and non-linear data, due to its advantage in data analysis separates the Support Vector Machine (SVM) Algorithm from all other models or algorithms therefore, Support Vector Machine(SVM) can be consider as proved of most preferable method to be used in case complex datasets [24-25].

Table 9: Compression of Different Machine Learning Model

SR.	Machine Learning Algorithms	Accuracy	Recall	Precision	F1-Score
01	Support Vector Machine (SVM)	97.5%	97.01%	97.8%	97%
02	Decision Tree (DT)	95.10%	96%	96.8%	95.18%
03	K-Nearest Neighbor(KNN)	95.18%	94.5%	94%	94%
04	Deep Learning Using Artificial Neural Network(DLANN)	98.28%	98%	98%	98%
05	Random Forest (RF)	97.2%	97.8%	97.6%	97%
06	Logistics Regression (LG)	95.10%	95.02%	96.5%	95.6%

CONCLUSION

While researching the respective topic the goal was to compare the working of several different algorithm models of machine learning. The dataset provided by the Wisconsin Breast Cancer (WBCD) was used in this research. The tumor cells (Malignant and Benign) were classified by using altered ML and DP algorithms. The result obtained from these models showed that the Adam gradient earned gas the highest accuracy on combining with RMSprop. The AdGrad is more suited for solving the problem or issues related to computer vision whereas the RMSprop is only used when there are non-stationary signals. Rectified Linear Unit (ReLU) was also used during the research to enable the fast learning of the model and to enhance their performance, the research can be also used as the base in the future prediction of breast cancer by means of imaging; as this is an important test or method for diagnosing cancer in patients.

Corresponding Author's Profile: The first Author Prof Dr Aftab Ahmad Malik PhD (University of Kent, Canterbury England); M.Phil; Msc; LL.B worked formerly as Chairman Department of Computer Science University of Engineering and Technology Lahore, Dean GIK Institute of Computer Science and Engineering KPK and Chairman & Director Department of Computer Science GC University Lahore as well as Dean University Education Lahore.

REFERENCES

[1] P. &. G. S. Gupta, "Breast cancer prediction using varying parameters of machine learning models. Procedia Computer Science," 2020.

[2] R. B. Prusty, "Increasing breast cancer awareness and breast examination practices among women through health education and capacity building of primary healthcare providers: a pre-post intervention study in a low socioeconomic area of Mumbai," 2021.

[3] U. Poehla, "awareness of breast cancer incidence and risk factors among healthy women in Germany: an update after 10 years. European Journal of Cancer Prevention," 2019.

[4] T. Takeshita, "Late recurrence of breast cancer is associated with the pro-cancerous immune microenvironment in the primary tumor," 2019; Scientific Reports 9(1):16942 ; DOI:10.1038/s41598-019-53482-x

[5] R. Tripathi, "Relative Examination of Breast Malignant Growth Analysis Utilizing Different Machine Learning Algorithms. International Journal of Perform Ability Engineering," 2022.

[6] P. Gupta, "Breast cancer prediction using varying parameters of machine learning models". Procedia Computer Science," 2020.

[7] Omondiagbe, D. A., Veeramani, S., & Sidhu, A. S. (2019, June). Machine learning classification techniques for breast cancer diagnosis. In IOP Conference Series: Materials Science and Engineering (Vol. 495, p. 012033). IOP Publishing.

[8] Asri, H., Mousannif, H., Al Moatassime, H., & Noel, T. (2016). "Using machine learning algorithms for breast cancer risk prediction and diagnosis". Procedia Computer Science, 83, 1064-1069.

[9] Agarap, A. F. M. (2018, February). On breast cancer detection: an application of machine learning algorithms on the wisconsin diagnostic dataset. In Proceedings of the 2nd international conference on machine learning and soft computing (pp. 5-9).

[10] Mohebian, M. R., Marateb, H. R., Mansourian, M., Mañanas, M. A., & Mokarian, F. (2017). A hybrid computer-aided-diagnosis system for prediction of breast cancer recurrence (HPBCR) using optimized ensemble learning. Computational and structural biotechnology journal, 15, 75-85.

[11] Doi, K. (2007). Computer-aided diagnosis in medical imaging: historical review, current status and future potential. Computerized medical imaging and graphics, 31(4-5), 198-211.

[12] R. Rangayyan, "A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs," Journal of the Franklin Institute Volume 344, Issues 3–4, May–July 2007, Pages 312-348.

[13] Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. Computational and structural biotechnology journal, 13, 8-17.

[14] Gayathri, B. M., Sumathi, C. P., & Santhanam, T. (2013). Breast cancer diagnosis using machine learning algorithms-a survey. International Journal of Distributed and Parallel Systems, 4(3), 105.

[15] Yi, L., & Yi, W. (2017, December). Decision tree model in the diagnosis of breast cancer. In 2017 International conference on computer technology, electronics and communication (ICCTEC) (pp. 176-179). IEEE.

[16] Nawi, N. M., Khan, A., Rehman, M. Z., Chiroma, H., & Herawan, T. (2015). Weight optimization in recurrent neural networks with hybrid metaheuristic Cuckoo search techniques for data classification. Mathematical Problems in Engineering, 2015.

[17] Gou, J., Du, L., Zhang, Y., & Xiong, T. (2012). A new distance-weighted k-nearest neighbor classifier. J. Inf. Computer. Science, 9(6), 1429-1436.

[18] Al-Shargabi, B., Alshami, F., & Alkhawaldeh, R. (2019). Enhancing multi-layer perception for breast cancer prediction. International Journal of Advanced Science and Technology.

[19] Mihaylov, I., Nisheva, M., & Vassilev, D. (2019). Application of machine learning models for survival prognosis in breast cancer studies. Information, 10(3), 93.

[20] David A. Omondiagbe, Shanmugam,Veeramani & Amandeep S. Sidhu 2019. "Machine Learning Classification Technique for Breast Cancer Diagnosis," Volume 495, 11th Curtin University Technology, Science and Engineering (CUTSE) International Conference 26–28 Sarawak, Malaysia *Conf. Ser.: Mater. Sci. Eng.* 495 012033; **DOI** 10.1088/1757-899X/495/1/012033

[21] Hajiabadi, H., Babaiyan, V., Zabihzadeh, D., & Hajiabadi, M. (2020). Combination of loss functions for robust breast cancer prediction. Computers & Electrical Engineering, 84, 106624.

[22] Jansen, T., Geleijnse, G., Van Maaren, M., Hendriks, M. P., Ten Teije, A., & Moncada-Torres, A. (2020). Machine learning explainability in breast cancer survival. In Digital Personalized Health and Medicine (pp. 307-311). IOS Press.

[23] Liu, H. X., Zhang, R. S., Luan, F., Yao, X. J., Liu, M. C., Hu, Z. D., & Fan, B. T. (2003). Diagnosing breast cancer based on support vector machines. Journal of Chemical Information and Computer Sciences, 43(3), 900-907.

[24] Rathi, M., & Pareek, V. (2016). Hybrid approach to predict breast cancer using machine learning techniques. International Journal of Computer Science Engineering, 5(3), 125-136.

[25] Swetha, K., & Ranjana, R. (2020). Breast cancer predication using machine learning and data mining. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 6(3), 610-5.

[26] Senapati, M. R., Mohanty, A. K., Dash, S., & Dash, P. K. (2013). "Local linear wavelet neural network for breast cancer recognition". Neural Computing and Applications, 22, 125-131.

[27]