PREDICTING OF BREAST CANCER DISEASE USING MACHINE LEARNING MODELS, A REVIEW

Oqbah Salim Atiyah
oqbah_salim@tu.edu.iq

Abstract

The breast cancer is a common disease in females, which causes death in women these days. Every year, an increase in cancer rates is recorded by about ٤,٥-٥٪ in all countries of the world. The early detection of breast cancer leads to an increase in the survival of those affected through diagnosis with mammography to provide appropriate treatment at the right time. With an X-ray, the breast is imaged by a radiologist, to be diagnosed and monitored by an oncologist to provide further treatment. Delay in diagnosing breast cancer leads to the spread of the disease in the body and thus reduces the chances of those affected surviving. Thus, utilizing machine learning to detect and diagnose breast cancer makes the process of identifying the disease more effective. This does not mean dispensing with the doctor, but rather helping him to understand the condition better and detect it early. This paper provides a summary of disease detection of breast cancer in those affected using machine learning techniques to increase the effectiveness and efficiency of prediction. Accuracy of disease diagnosis and accuracy of classification are the main objectives.

Keywords: Breast cancer datasets, Machine Learning, Patient features.

١. INTRODUCTION

Breast cancer is a disease that spreads among females all over the world and is the most common cancer. It has become the leading cause of cancer-related deaths among females [١]. In recent years, it has been observed that the incidence of breast cancer in females has increased, as announced by the World Health Organization, [٢] where there are about ١,٣٨ million new infections and about ٤٥٨,٠٠٠ deaths every year as a result of breast cancer. This disease is most prevalent in developing and developed countries. Mortality rates for breast cancer, the most common type, exceed those of other types of cancer except for lung cancer [٣]. Survival rates for

people with breast cancer vary greatly from one place to another around the world, ranging from ٤٠٪ in low-income countries, about ٦٠٪ in middle-income countries, and ٨٠٪ or more in high-income countries, due to the use of machine learning techniques and modern devices in medical diagnosis [٤]. Breast cancer is more aggressive than other cancers. Factors that lead to breast cancer are ionizing radiation, sex, obesity, alcohol consumption, lack of exercise, advanced age, and hormonal imbalance, especially in the premenstrual period. These above effects are not the only ones, there are other factors in the genetic makeup that cause breast cancer, its spread, and increase its aggressiveness [٥]. Therefore, studies applied in the medical field related to breast cancer are one of the main things to provide effective diagnostic systems to predict the disease with high accuracy to provide appropriate treatment for those affected at the appropriate time. Advanced methods of breast cancer screening allow healthcare professionals to diagnose breast cancer early. Because detecting cancer early increases the hope of treating it. Even if it cannot be cured, there are many treatments to prolong life. Discoveries in breast cancer research help healthcare professionals choose the most effective treatment plans. There is no doubt that evaluating the information taken from infected people and the decisions of specialists are the most important in detection and diagnosis [٦]. Still, the use of artificial intelligence systems and specialized systems greatly helps experts in detecting and predicting the disease. With the help of these systems, possible errors that may be made by medical professionals during diagnosis can be avoided. Also, the process of examining medical data is more detailed and takes a shorter time [٧]. In the medical field, classification models and data mining and analysis methods are widely used to aid detection, diagnosis, and decision-making.

Fig. ١. Display machine learning roles in the detection of breast cancer.

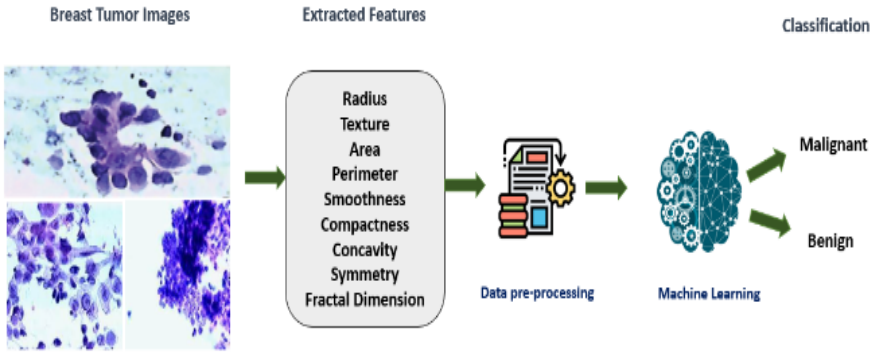

Fig ١. Display ML in breast cancer detection [٣٦].

٢. MACHUNE LEARNING

It is a subpart of artificial intelligence, which trains and teaches computers from data rather than programming them clearly and explicitly to do tasks correctly. It includes two phases: in the training phase, the techniques are trained to detect patterns and identify correlations in data sets, and in the testing phase, decisions and predictions are made based on this analysis. Machine training is developed and improved with use to become better and more accurate the larger the data that can be accessed [٨].

Statistical models and strategies in machine learning are trained and learned to accomplish their tasks without being explicitly instructed. As the devices work and learn automatically. Machine learning aims to develop computer methods through training and learning processes to access data [٩]. Machine learning has been applied in various areas of life such as commerce, natural disasters, healthcare, etc. [١٠]. ML is important in discovering and determining training results based on the principle of taking a portion of the data set. Undefined dependencies of the model in the provided dataset are identified. Prediction results are the new outputs of whether the dependencies specified in the model are known or not. There is many machine learning methods introduced in biomedical research in healthcare. Machine learning uses various methods and models to generalize data across broad a data set. Machine learning trains and learns in a variety of ways, as in Fig ٢.

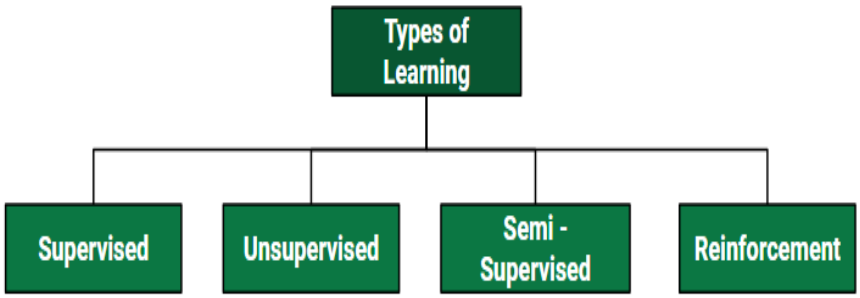

Fig ٢. Shows types learn in machine learning [٥٢].

٢,١ SUPERVISED LEARNING General Consideration

The learning in this type is based on categorical data. The algorithm must learn how to summarize the nature of the training data acceptably by following training strategies. The data is divided into two parts: the largest part for training and the other for testing. The chosen model is applied to the training set after which its performance is tested through the use of performance metrics [١١]. Supervised learning solves issues related to classification and regression. The categorical value is separate in the classification. While the values are continuous in the regression process. Classification models are used to predict objects that belong to classes in a data set, while regression models work to predict objects when the data is continuous [١٢].

٢,٢ UNSUPERVISED LEARNING

The learning in this type relies on datasets with unlabeled inputs. Cluster analysis is more used in this type of learning than other models, as it works to integrate hidden information or data or explore patterns in groups. In unsupervised learning, the model is applied to the training data while the values of the target classes must be unknown [١٠]. The model groups similar data by discovering and identifying hidden patterns. In this type of learning, the performance of the user model cannot be evaluated because the value of the categories is unknown [١٣].

٢,٣ REINFORCEMENT LEARNING

The learning in this type relies on labeled and unlabeled data. It is a combination of supervised learning and unsupervised learning that is based on categorical and non-categorical data. Unsorted data must be longer than sorted data. The model's performance is highly efficient when both categorical and non-categorical data are present in the data set [١٤].

٢,٤ REINFORCEMENT LEARNING

The learning in this type depends on learning new acceptable strategies and patterns based on environmental awareness behavior. Learning models perform influences on the environment to create inputs that help them make predictions. Technologies work to improve and develop their performance through interaction based on the environment without human intervention [١٥].

A. K-NEAREST NEIGHBOR(KNN)

It is a technique that classifies data by creating a new sample that is attached to existing samples within a specific range and for samples with similar features, a nearest neighbor is calculated. The value of k represents a user-specific number. Each specific sample finds similar features in the new sample, and this leads to classifying the new samples into similar categories. [١٦, ١٧].

B. SUPPORT VECTOR MACHINE (SVM)

It is a technique that solves issues related to regression and classification. It is used to classify input data with categorical samples. This algorithm works by dividing the data set into two parts and classifies categorical data by measuring the farthest margin values between the hyperlevel values and the closest data point to determine its belonging to any categorical part of the data sets. [١٨,١٦,١٩,٢٠,٢١,٢٢,٢٣].

C. DECISION TREE

It is a technique that solves issues of both regression and classification and can be applied to continuous and categorical features. The shape of a decision tree is similar to a tree. From the top, it begins with a root node. It then branches into several nodes to form many branches that represent solutions. This algorithm

provides a solution through documented steps for each process to achieve the desired goal. The possible outputs of the selected samples are the most specific results. The model works to predict the indicated outputs to determine and test the results. [٢٤, ١٩, ٢٥, ١٧, ٢٦].

E. NAÏVE BAYES CLASSIFIER

It is an algorithm based on probability and statistics. These techniques work based on variable scores to make filtering decisions from different samples and can be applied to large data sets. The main goal of these classifiers is to predict the chosen class and its belonging to the specified group. Bayesian classifier models are characterized by independence between predictors. [١٧, ٢١].

F. RANDOM FOREST

The random forest technique takes the same approach as a decision tree. A random forest consists of several decision trees. These trees are recreated by randomly selecting data. The algorithm chooses a set of samples to optimize each of the various decision trees. The final decision is made by determining the highest number of vote outputs from the multiple trees. An ensemble classifier is used. [٢١, ٢٧].

G. K-MEAN ALGORITHM

Clustering is a common method in unsupervised learning [٢٨] that divides a data set into clusters, k is a predefined number, representing the iterative method of numbering clusters. K-Means clustering techniques examine the relevance rate of outputs and determine similarities between outputs by searching and clustering web pages such as Yahoo, Google, etc.

٣. LITERATURE SURVEY

The researchers in this paper [١٨] used several SVM techniques and SVM ensembles on breast cancer datasets. The SVM-based boosting and RBF kernel method were used to achieve the highest prediction accuracy, and evaluation parameters such as F-measure and ROC curve were calculated to create a strong

model. The researchers discovered that the SVM-based boosting and RBF kernel method give higher results than other models.

In the paper [١٦], researchers used a new approach based on high-dimensional feature identification of microarrays to classify cancer data. New filter models are used to develop the signal-to-noise ratio. They found that PSO achieves the best results when used with SVM, k-NN, and PNN. The breast cancer dataset includes ٩٧ cases containing about ٢٤,٤٨١ genes. The new approach applied to the breast cancer dataset achieved results with ١٠٠٪ accuracy.

The researchers in this study [٢٠] tried to find a solution to the impression syndrome in breast cancer datasets that has advanced and improved by the presented research studies and the presence of social networking sites using the K-medoid clustering method. The K-medoid clustering approach has also been improved, which has improved the clustering efficiency by reimaging the breast cancer syndrome and presenting the image with a moderate negative silhouette to other selected clusters after the initial implementation of K-medoid clustering.

In this paper [١٧] researchers compared different ML techniques: C٤,٥ SVM, kNN, and NB for the same breast cancer dataset of about ٦٩٩ cases and ١١ features found in the WBCD website. The result of the SVM algorithm was ٩٧,١٣٪, which is the highest accuracy among all the algorithms used, and with a lower error rate than other mining models implemented on the data on the WEKA website.

In this paper [٢١] the authors used three machine learning techniques: Naïve Bayes, SVM, and RF on a breast cancer dataset. The RF outputs were the most suitable with an accuracy of about ٩٩,٤٢٪, while the NB and SVM results were ٩٨,٢٤٪ and ٩٨,٨٪, respectively.

In this paper [٢٩], the researchers used a hybrid system based on neural networks and association rule definition models. This approach is based on evolutionary techniques in which the amplitude problem is used to diagnose breast cancer. The development of the arm is based on optimizing the selection of the most important

features to reduce the amplitude scales to find the correlation between multiple samples, after which the NN technique is used for active and effective distribution. In the paper [٣٠]. The authors used a powerful new computational method to determine gene identity. Initially, a traditional t-test technique was used to reduce scaling in the process of testing the data set. The new approach was then used to find important genetic codes by developing and improving specific particle assemblies, the chosen technique that contains SRBCT data, thus initializing the subassemblies into other groups.

In this paper [٣١] the focus was on data-finding models to identify patterns and their relationships. Machine learning techniques were used to diagnose breast cancer and predict recurrence in patients who were followed up by the Iranian Breast Cancer Center. The techniques used are Support Vector Machine (SVM), Decision (DT), and Artificial Neural Network (ANN). SVM achieved a higher result than the other, with an accuracy of ٩٥,٧٪.

In this study [٣٢] the authors used machine learning models on two breast cancer datasets (BCCD and WBCD). The performance of the techniques used such as Random Forest (RF), DT, SVM, and Logistic Regression (LR) was measured. achieved SVM has the highest accuracy among the algorithms performed on both datasets (BCCD and WBCD), with an accuracy of ٩٧,٧ and ٧٦,٣, respectively.

In this paper [٣٣], the authors used several machine learning techniques in medical image processing for breast cancer screening and diagnosis on a dataset that includes X-ray images performed on patients. The authors found that the SVM algorithm gives the best result with an accuracy of ٩٤٪.

In this study [٣٤]. Researchers use random forest techniques and feature selection methods by identifying useful features in the data set and discarding their extra features. The method achieved ١٠٠٪ classification accuracy, which is a very promising result.

Researchers were used in this paper [٣٥]. Support vector machine and rough set (RS_SVM) on a dataset, and this method achieved good results with an accuracy of ٩٦,٨٧٪.

٤. COMPARISON OF MACHINE LEARNING

With the continuous developments in specialized systems and machine learning techniques and their entry into various areas of life, including the medical field. Decision-making in the medical field may sometimes cause some problems. Using these advanced systems in this field facilitates the detection process in time and errorless. During the reviews, we found that several types of cancer are detected and predicted in several ways. The characteristics and behavior of each disease vary depending on the type of cancer. One of these diseases is breast cancer which affects females. Modern techniques were used on data sets obtained from various sources. Results achieved from machine learning strategies are very important in detecting and diagnosing disease.

We have classified the machine learning techniques and breast cancer datasets as shown in the table below. Table ١ presents the classification techniques implemented in several research papers for breast cancer detection and prediction. This table displays various datasets that were obtained from various sources to train machine learning techniques on and measure their accuracy. In our review of studies, we noted that the machine learning algorithms used to diagnose and predict breast cancer were implemented after the dataset was collected. Next, preprocess the data to identify unimportant features, and reduce the dimensions of the data because using the entire features as in the data set reduces the accuracy of the results.

Table ١. Show the accuracy measured of ML models used of breast cancer.

S.No.	Author	Algorithm/method	Dataset	Accuracy
١-	[١٦] S. Barnali	PSO-KNN PSO-SVM	GEDatasets	١٠٠٪
٢-	[١٧] A. Hiba,	C٤,٥ SVM	WBCD	٩٥,١٣ ٩٧,١٣

		NB k-NN		٩٥,٩٩ ٩٥,٢٧
٣-	[٣٧] Ripon Patgiri, et al.	SVM ANN	WBCD	٩٩,٤١٪ ٨٧,١٣٪
٤-	[٣٨] Yixuan LI, et al.	SVM	BCCD WBCD	٩٧,٧٪ ٧٦,٣٪
٥-	[٣٩] Yun Liu, et al.	CNN	CAMELYON١٦ challenge	٩٧٪
٦-	[٣٠] A. Alhasani, et al.	Decision trees (DT) SVM NB	WDBC	١٠٠٪ ٩٨,٤٢٪ ٦٩,٦٠٪
٧-	[٤٠] Neslihan Bayramoglu	CNN	BreakHis	٨٣,٢٥٪
٨-	[٤١] Daniel Lévy	CNN	DDSM	٩٢,٣٪
٩-	[٤٢] Muhammad Talha	Improvement mammograms, fusion in discrete wavelet transform, and discrete cosine transform with SVM classifier	MIAS	٩٦٪
١٠-	[٤٣] Mellisa Pratiwi	classification based on gray level co-occurance matrix (GLCM) texture based features	MIAS	٩٣,٩٨ %.
١١-	[٤٤] J. Dheeba	POSWNN	real clinical data	٩٦,٨٪
١٢-	[٤٥] M. Muthu Rama Krishnan	Statistical analysis of mammographic features combined with support vector machine	WBCDD	٩٣,٧٢٦
١٣-	[٤٦] uxandra Stoean	Support vector machines and evolutionary algorithms	WBCDD WBCPD	٩٧ ٧٩
١٤-	[٤٧] Zhiwang Zhang	Rough set-based multiple criteria linear programming approach	WBCPD WBCDD	٦٥ ٨٩
١٥-	[٣٤] Nguyen, C.	Support vector machine and rough set	WBCPD WBCDD	٩٩,٧ ٩٩,٨٢
١٦-	[٤٨] Tingting Mu	Support vector machines, radial basis function networks and self-organizing maps	WBCDD	٩٨
١٧-	[٤٩] Der-Chiang Li	A fuzzy-based data transformation for feature extraction to increase classification performance	WBCDD	٩٦,٣٥
١٨-	[٥٠] Sean N. Ghazavi	Three fuzzy modeling methods including the fuzzy	WBCDD	٩٧,١٧

		k-nearest neighbor algorithm		
١٩-	**[٥١] J. Wu**	**K-nearest neighbor (kNN)** **Naïve Bayes (NGB)** **Decision trees (DT)** **Support Vector Machines (SVM)**	**Genomics Data Commons (GDC)**	٨٧٪ ٨٥٪ ٨٧٪ ٩٠٪

٥. CONSTRUCTION

The review of effective models for breast cancer disease classification and prediction was presented through the research process in the field of biomedical informatics. From this extensive review, we discovered that many automatic breast cancer prediction methods rely on machine learning techniques including clustering and classification strategies. This survey provided an extensive review of several machine learning concepts and classification strategies applied to several types of breast cancer datasets for prediction and detection.

The findings that researchers have achieved to overcome challenges in breast cancer diagnosis using various machine learning models were tabulated and presented in detail. The most accurate and superior techniques are SVM and decision trees, which achieved the highest results on the various types of data sets used. On larger datasets, the accuracy decreased slightly. Therefore, dimensionality reduction strategies should be used to select the most important features. There is still a need to develop a prediction of early-stage breast cancer. There are many established data sets that help in searching for more information and knowledge about breast cancer.

REFERENCES

[١] https://doi.org/١٠,٥٤٢١٦/JAIM.٠٤٠٢٠١

[٢] WHO (٢٠١٢) Breast cancer: Prevention and control. http://www.who.int/cancer/detection/breastcancer/en/indexx١.html

[٣] The Women's Health Resource (٢٠١٣) What is breast cancer. http://www.imaginis.com/general-information-on-breastcancer/what-is-breast-cancer-٢

[٤] Béatrice Lauby Secretan et al.," Breast-Cancer Screening — Viewpoint of the IARC Working Group", The New England Journal of Medicine, n engl j med ٣٧٢;٢٤ nejm.org June ١١, ٢٠١٥, p.p. ٢٣٥٣-٢٣٥٨

[٥] Ince, M.C. and Karabatak, M. (٢٠٠٩) An expert system for detection of breast cancer based on association rules and neural network. *Expert Systems with Applications,* ٣٦, ٣٤٦٥-٣٤٦٩. doi: ١٠,١٠١٦/j.eswa.٢٠٠٨,٠٢,٠٦٤

[٦] Sahana, S., Polat, K., Kodaz, H. and Günes, S. (٢٠٠٧) A new hybridmethod based on fuzzy-artificial immune system and k-nn algorithmfor breast cancer diagnosis. *Computers in Biology and Medicine,* ٣٧٧, ٤١٥-٤٢٣. doi:١٠,١٠١٦/j.compbiomed.٢٠٠٦,٠٥,٠٠٣

[٧] Min-Wei Huang, Chin-Wen Chen, Wei-Chao Lin, Shih-Wen Ke, Chin-Fong Tsai, "SVM and SVM Ensemble in Breast Cancer Prediction ", PloS one vol.١٢, no.١, ٢٠١٧. DOI: ١٠,١٣٧١/journal.pone.٠١٦١٥٠١, January ٦, ٢٠١٧.

[٨] A. Abdullah Alfayez, H. Kunz and A. Grace Lai. Predicting the risk of cancer in adults using supervised machine learning: a scoping review. BMJ Open ٢٠٢١ Vol. ١١ Issue ٩ Pages e٠٤٧٧٥٥ DOI: ١٠,١١٣٦/bmjopen-٢٠٢٠-٠٤٧٧٥٥

[٩] Yu-Xiong Wang, Martial Hebert, "Learning to Learn: Model Regression Networks for Easy Small Learning", European Conference on Computer Vision (ECCV ٢٠١٦) Lecture Notes in Computer Science, vol.٩٩١٠, pp ٦١٦-٦٣٤, ٢٠١٦ .DOI:١٠,١٠٠٧/٩٧٨-٣-٣١٩-٤٦٤٦٦-٤_٣٧

[١٠] E. Alpaydin. "Introduction to Machine Learning". MIT press; ٢٠٠٩.

[١١] T. Yuchun, Yan-Qing Zhang, Zhen Huang, Xiaohua Hu, Yichuan Zhao, "Recursive Fuzzy Granulation for Gene Subsets Extraction and CancerClassification", IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE, vol. ١٢, no. ٦, pp. ٧٢٣-٧٣٠, NOVEMBER ٢٠٠٨.

[١٢] D. Jayashree, Sanjit K Dash, Swet Dash, Madhusmita Swain, "A Classification Technique for Microarray Gene Expression Data using PSOFLANN", International Journal on Computer Science and Engineering, vol. ٤, no.٩, pp. ١٥٣٤-١٥٣٩. Sep ٢٠١٢,

[١٣] S. Akanksha , Parminder Kaur, "Optimized Liver Tumor Detection and Segmentation Using Neural Network", International Journal of Recent Technology and Engineering (IJRTE), vol. ٢, no. ٥, , pp ٧-١٠, November ٢٠١٣.

[١٤] Y. Hu, K. Ashenayi, R. Veltri, G. O'Dowd, G. Miller, R. Hurst and R. Bonner, "A Comparison of Neural Network and Fuzzy c-Means Methods in Bladder Cancer Cell Classification", Proceedings of ١٩٩٤ IEEE International Conference on Neural Networks (ICNN'٩٤), pp ٣٤٦١- ٣٤٦٦, ١٩٩٤. DOI: ١٠,١١٠٩/ICNN.١٩٩٤,٣٧٤٨٩١

[١٥] T.O. Ayodele, "Types of machine learning algorithms. New advances in machine learning", vol. ٣, pp. ١٩-٤٨. ٢٠١٠.

[١٦] S. Barnali, Debahuti Mishra, "A Novel Feature Selection Algorithm using Particle Swarm Optimization for Cancer Microarray Data‖", International Conference on Modeling Optimization and Computing (ICMOC-٢٠١٢), ELSEVIERProcedia Engineering, vol.٣٨, pp. ٢٧–٣١, ٢٠١٢.

[١٧] A. Hiba, Hajar Mousannifb, Hassan Moatassimec, Thomas Noeld, "Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis", ELSEVIER ٦th International Symposium on Frontiers in Ambient and Mobile Systems (FAMS ٢٠١٦), vol. ٨٣, pp ١٠٦٤–١٠٦, ٢٠١٦.

[١٨] Min-Wei Huang, Chin-Wen Chen, Wei-Chao Lin, Shih-Wen Ke, Chin-Fong Tsai, "SVM and SVM Ensemble in Breast Cancer Prediction ", PloS one vol.١٢, no.١, ٢٠١٧. DOI: ١٠,١٣٧١/journal.pone.٠١٦١٥٠١, January ٦, ٢٠١٧.

[١٩] K. Sivakami, "Mining Big Data: Breast Cancer Prediction using DT - SVM Hybrid Model", International Journal of Scientific Engineering and Applied Science (IJSEAS), vol. ١, no. ٥, pp. ٤١٨-٤٢٩, august ٢٠١٥.

[٢٠] Q. Ping, C. C. Yang, S. A. Marshall, N. E. Avis, and E. H. Ip, "Breast cancer symptom clusters derived from social media and research study data using improved K-Medoid clustering", in IEEE Transactions on Computational Social Systems", vol. ٣, no. ٢, pp. ٦٣–٧٤, June ٢٠١٦.

[٢١] B. Dai, Chen, R. C., Zhu, S. Z., & Zhang, W. W. "Using random forest algorithm for breast cancer diagnosis". In ٢٠١٨ International symposium on computer, consumer and control (IS٣C) IEEE. pp. ٤٤٩-٤٥٢, December ٢٠١٨.

[٢٢] D. Urun, Tobias Glasmachers, Christian Igel, "A Unified View on Multi-class Support Vector Classification", Journal of Machine Learning Research, vol. ١٧, pp. ١-٣٢, ٢٠١٦.

[٢٣] A. Mennatallah, Markus Goldstein, Slim Abdennadher, "Enhancing One-class Support Vector Machines for Unsupervised Anomaly Detection", In Proceedings of the ACM SIGKDD ٢٠١٣ Workshop on Outlier Detection and Description, pp. ٨-١٥. ٢٠١٣.

[٢٤] P. Ramachandran, N.Girija, T.Bhuvaneswari, "Early Detection and Prevention of Cancer using Data Mining Techniques", International Journal of Computer Applications, vol. ٩٧, no.١٣, pp. ٤٨-٥٣, July ٢٠١٤.

[٢٥] C. Vikas, Saurabh Pal, "Data Mining Techniques: To Predict and Resolve Breast Cancer Survivability", International Journal of Computer Science and Mobile Computing, vol.٣ no.١, pp. ١٠-٢٢ January, ٢٠١٤.

[٢٦] Ahmad LG, Eshlaghy AT, Poorebrahimi A, Ebrahimi M, Razavi AR, "Using Three Machine Learning Techniques for Predicting Breast Cancer Recurrence", Journal of Health & Medical Informatics, vol ٤, no. ٢, ٢٠١٣, http://dx.doi.org/١٠,٤١٧٢/٢١٥٧-٧٤٢٠,١٠٠٠١٢٤.

[٢٧] N. Cuong, Yong Wang, Ha Nam Nguyen," Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic", J. Biomedical Science and Engineering, vol. ٦, pp ٥٥١-٥٦٠, ٢٠١٣, DOI: http://dx.doi.org/١٠,٤٢٣٦/jbise.٢٠١٣,٦٥٠٧٠.

[٢٨] T. Arunanand, Abdul Nazeer K A, Mathew J P, Meeta Pradhant, "A Nature-inspired Hybrid Fuzzy C-means algorithm for Better Clustering of Biological Data Sets", IEEE International Conference on Data Science & Engineering (ICDSE _١٤), pp ٧٦-٨٢, ٢٠١٤.

[٢٩] S. Marshland, "Machine Learning an Algorithmic Perspective". CRC Press, Conference, Adelaide, Australia. Conferences in Research and Practice in Information Technology, vol. ١٩, ٢٠٠٣.

[٣٠] A. Alhasani, et al. "A comparative analysis of methods for detecting and diagnosing breast cancer based on data mining". Methods. vol.٧, no.٩, ٢٠٢٣.

[٣١] T.O. Ayodele, "Types of machine learning algorithms. New advances in machine learning", vol. ٣, pp. ١٩-٤٨. ٢٠١٠.

[٣٢] Y. Hu, K. Ashenayi, R. Veltri, G. O'Dowd, G. Miller, R. Hurst and R. Bonner, "A Comparison of Neural Network and Fuzzy c-Means Methods in Bladder Cancer Cell Classification", Proceedings of ١٩٩٤ IEEE International Conference on Neural Networks (ICNN'٩٤), pp ٣٤٦١- ٣٤٦٦, ١٩٩٤. DOI: ١٠,١١٠٩/ICNN.١٩٩٤,٣٧٤٨٩١

[٣٣] Miles N. Wernick et al.,"Machine Learning in Medical ImagingDrawing drawing conclusions from images", IEEE SIGNAL PROCESSING MAGAZINE, ١٠٥٣-٥٨٨٨/١٠/$٢٦,٠٠©٢٠١٠IEEE.

[٣٤] Nguyen, C., Wang, Y., & Nguyen, H. N. Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic. ٦, ٥٥١-٥٦٠, (٢٠١٣).

[٣٥[Keles, A., Keles, A. and Yavuz, U. (٢٠١١) Expert system based on neuro-fuzzy rules for diagnosis breast cancer. *Expert Systems with Applications,* ٣٨, ٥٧١٩-٥٧٢٦. doi:١٠,١٠١٦/j.eswa.٢٠١٠,١٠,٠٦١

[٣٦] Pic١_ T. wisdomml https://wisdomml.in/breast-cancer-prediction-using-machine-learning/ April ٣, ٢٠٢٣ by

[٣٧] Ripon Patgiri et al., (٢٠١٩,١٧ March)," Machine Learning: A Dark side of Cancer computing" International Conference. Bioinformatics and Computational Biology BIOCOMP' ١٨, ISBN: ١-٦٠١٣٢-٤٧١-٥, CSREA Press.

[٣٨] Yixuan LI, Zixuan Chen," Performance Evalaution of Machine Learning Methods for Breast Cancer Prediction," doi: ١٠,١١٦٤٨.j.acm.٢٠١٨٠٧٠٤,١٥.

[٣٩] Yun Liu et al.," Detecting Cancer Metastases on Gigapixel Pathology Images," arXiv:١٧٠٣,٠٢٤٤٢v[cs.CV] ٨ Mach ٢٠١٧.

[٤٠] Neslihan Bayramoglu et al.," Deep Learning for Magnification Independent Breast Cancer Histopathology Image Classification", ٢٠١٦ ٢٣rd International Conference on Pattern Recognition (ICPR) Cancún Center, Cancún, México, December ٤-٨, ٢٠١٦

[٤١] Daniel Lévy. (٢٠١٦, December) Breast Mass Classification from Mammograms using Deep Convolutional Neural Networks presented at ٣٠th Conference on Neural Information Processing Systems (NIPS ٢٠١٦), Barcelona, Spain

[٤٢] Muhammad Talha," Classification of mammograms for breast cancer detection using fusion of discrete cosine transform and discrete wavelet transform features", Biomedical Research ٢٠١٦; ٢٧ (٢): ٣٢٢-٣٢.

[٤٣] Mellisa Pratiwi et al.," Mammograms Classification using Gray-level Co-occurrence Matrix and Radial Basis Function Neural Network," Procedia Computer Science ٥٩ (٢٠١٥) ٨٣ – ٩١.

[٤٤] Dheeba, V., Albert Singh, N., & Amar Pratap Singh, J. (٢٠١٤). Breast cancer diagnosis: an intelligent detection system using wavelet neural network. In *Proceedings of the international conference on frontiers of intelligent computing: theory and applications (FICTA) 2013* (pp. ١١١-١١٨). Springer International Publishing.

[٤٥] Krishnan, M.M.R., Banerjee, S., Chakraborty, C., Cha kraborty, C. and Ray, A.K. (٢٠١٠) Statistical analysis of

mammographic features and its classification using support vector machine. ***Expert Systems with Applications,*** **٣٧, ٤٧٠-٤٧٨. doi: ١٠,١٠١٦/j.eswa.٢٠٠٩,٠٥,٠٤٥**

[٤٦] Stoean, R. and Stoean, C. (٢٠١٣) Modeling medical decision making by support vector machines, explaining by rules of evolutionary algorithms with feature selection. ***Expert Systems with Applications,*** **٤٠, ٢٦٧٧-٢٦٨٦. doi: ١٠,١٠١٦/j.eswa.٢٠١٢,١١,٠٠٧**

[٤٧] Zhang, Z.W., Shi, Y. and Gao, G.X. (٢٠٠٩) A rough setbased multiple criteria linear programming approach for the medical diagnosis and prognosis. ***Expert Systems with Applications,*** **٣٦, ٨٩٣٢-٨٩٣٧. doi: ١٠,١٠١٦/j.eswa.٢٠٠٨,١١,٠٠٧.**

[٤٨] Mu, T.T. and Nandi, A.K. (٢٠٠٧) Breast cancer detection from FNA using SVM with different parameter tuning systems and SOM-RBF classifier. ***Journal of the Franklin Institute,*** **٣٤٤, ٢٨٥-٣١١. doi: ١٠,١٠١٦/j.jfranklin.٢٠٠٦,٠٩,٠٠٥**

[٤٩] Li, D.-C., Liu, C.-W. and Hu, S.C. (٢٠١١) A fuzzy-based data transformation for feature extraction to increase classification performance with small medical data sets. ***Artificial Intelligence in Medicine,*** **٥٢, ٤٥-٥٢. doi: ١٠,١٠١٦/j.artmed.٢٠١١,٠٢,٠٠١**

[٥٠] Ghazavi, S.N. and Liao, T.W. (٢٠٠٨) Medical data mining by fuzzy modeling with selected features. ***Artificial Intelligence in Medicine,*** **٤٣, ١٩٥-٢٠٦.**

[٥١] J. Wu, and C. Hicks. "Breast cancer type classification using machine learning". Journal of personalized medicine. vol.١١, no.٢, p. ٦١, ٢٠٢١.

[٥٢]https://www.geeksforgeeks.org/semi-supervised-classification-in-data-mining/
Updated ٠٨ Aug, ٢٠٢٢