An Intelligent System for Predicting the Breast Cancer Threat Using Health Data Registry and Awareness: A Review

Tamil Selvi Madeswaran, Aruna Kumar Kavuru, Padma Theagarajan, Nasser Al Hadhrami, Maya Al Foori, and Ohm Rambabu

***Abstract* — According to the World Health Organization, there are around 2.1 million new cases of breast cancer each year, making it the most often diagnosed cancer in women that is serious. Omani women reported having breast cancer at a rate of roughly 34.1% overall, with an average age of 34.7 and high fatality rates of 11 per 100,000 people (GLOBOCAN 2018). Changes in lifestyle and risk factors include age, family history, early menopause, late menopause, obesity, and birth control pills are the main causes of breast cancer. Recent research observations shown that a combination of risk variables is responsible for the prevalence of breast cancer. Sometimes unidentified risk factors will also contribute to the development of breast cancer. Also, the influence of how much each risk factor contributes to the development of cancer differs across females. The objective of this study is to review supervised machine learning methods, including Logistic Regression, Neural Networks, Decision Trees, and Nearest Neighbors, in order to forecast the likelihood of breast cancer occurring among females.**

***Keywords* — Artificial Intelligence, Breast Cancer, Machine Learning, Prediction, Regression, Risk Factors.**

I. Introduction

Cancer patients experience difficult physical problems. Cancer diagnosis results in significant emotional pain, and its treatments have harmful side effects. Many cancer patients view their initial diagnosis as a catastrophic event that will leave them depressed and anxious. The effects of cancer are also upsetting for the family. The difficult work of cancer control and prevention is a crucial component of the healthcare system. Its goal is to improve the quality of life for cancer patients while reducing the incidence, mortality, and morbidity of care. In order to achieve this goal, accurate, sufficient, full, and timely health information and clinical knowledge are essential. The suggested research intends to review the Knowledge-based Prediction Model to address such a life-threatening condition.

The goal of the proposed study is to examine the knowledge-based prediction model that predicts the likelihood of breast cancer among the female population.

These days various supervised Machine Learning algorithms are used for detecting breast cancer since the ML produce the accuracy and prediction in the chances of recurrence.

In this paper we have discussed various ML algorithms which can be used for Breast Cancer prediction like, Logistic Regression, Neural Networks, Decision Trees and Nearest Neighbors etc. The accuracy, precision and recall of all the algorithms are compared to find out which algorithm is most suited for Breast Cancer detection.

The rest of the paper is organized as follows: Section II discusses the Breast cancer, types and stages. Section III discusses the several methods for cancer classification and prediction. Section IV describes the hybrid approach for prediction and classification. Section V concludes the paper.

II. Stages of Breast Cancer

A. Breast Cancer

Middle-aged women's morality has significant roots in breast cancer. It can be decreased by early detection of breast cancer. It is one of the most prevalent cancer types and the leading cause of mortality for women around the world.

There are different types and subtypes of breast cancer. Treatments for the breast cancer is quite different from other treatments that women may have. Main type of breast cancer is Non-invasive and Invasive which represents spread of cancer in the breast. Non-invasive type of cancer spreads the cancer within the milk ducts or lobules in the breast. In contrast, invasive breast cancer grows into nearby breast tissue from the ducts or lobules of the breast.

The majority (90%) of non-invasive breast cancer cases are caused by ductal carcinoma in situ (DCIS). Lobular carcinoma in situ (LCIS) is regarded as a higher risk factor for developing breast cancer [1].

There were various research papers published where the stages of breast cancer are classified and, in this paper, these stages are described, classified based on the various algorithms used.

Submitted on March 01, 2023.
Published on May 06, 2023.
T. S. Madeswaran, Dr., University of Technology and Applied Sciences, Nizwa, Oman.
(e-mail: e608026@utas.edu.om)
A. K. Kavuru, University of Technology and Applied Sciences, Nizwa Oman.
(e-mail: e606015@utas.edu.om)
P. Theagarajan, Sona College of Technology, India.
(e-mail: padmatheagarajan@gmail.com)
N. Al Hadhrami, University of Technology and Applied Sciences, Nizwa, Oman.
(e-mail: Nasser.almur@nct.edu.om)
M. Al Foori, Royal Hospital, Oman.
(e-mail: lianobaby@gmail.com)
O. Rambabu, University of Technology and Applied Sciences, Nizwa, Oman.
(e-mail: ram.babu@nct.edu.om)

B. Stages of Breast Cancer

Sun *et al.* [2] has classified the various stages of breast cancer in their research. They are classified into their major areas with multiple stages in each area and are shown in Table I, Table II, and Table III.

TABLE I: IN SITU CARCINOMA

Area 1: In situ Carcinoma [2]	
Stage	Description
O	This is the stage where the tumor is confined, to a milk duct or a milk producing gland. This has not entered nearby breast tissue. [2]

TABLE II: LOCALIZED AND REGIONAL INVASIVE CANCER

Area 2: Localized and regional invasive Cancer [2]	
Stage	Description
I	A tumor with ¾ inch or 2 cm diameter, but not yet entered or spread beyond the breast. [2]
IIA	In this stage the tumor is ¾ inch or less in diameter but has spread one to three lymph nodes in the armpit. [2]
IIB	the tumor is larger than ¾ inch, smaller than 2 inch in diameter and has not spread beyond the breast. [2]
IIIA	The tumor is 2 inches or less or might have spread to four to nine lymph nodes in the armpit or near breastbone.[2]
IIIB	The tumor is spread to the chest wall or skin.[2]
IIIC	The tumor can be of any size and might have spread any or many of the various places near breast or armpit. [2]

TABLE III: MATASTATIC CANCER

Area 3: Metastatic Cancer [2]	
Stage	Description
IV	The tumor regardless of size has spread to many organs such as lungs or bones etc., [2]

III. SIGNIFICANT MACHINE LEARNING ALGORITHMS IN PREDICTING BREAST CANCER

During the literature review it was noticed that there exists a huge amount of research already carried out in this area with various algorithms that are significant enough in predicting the breast cancer. The following are some of the most commonly used algorithms explain in brief.

A. Naive Bayesian Classifier

The Naive Bayesian classifier is a machine learning algorithm that uses Bayes' theorem to predict the probability of breast cancer based on past data that consists of a set of features or variables (such as age, family history, mammogram results, etc.) to calculate the likelihood of a patient having breast cancer. The significance of the Naive Bayesian classifier in breast cancer prediction lies in its ability to make predictions quickly and accurately. It is considered one of the simplest and most straightforward algorithms, making it accessible and easy to use for medical professionals. Additionally, the Naive Bayesian classifier can handle high-dimensional datasets, making it suitable for large-scale breast cancer screening programs. It also has the ability to deal with missing or incomplete data, which is common in medical datasets. Overall, the use of the Naive Bayesian classifier in breast cancer prediction can lead to earlier and more accurate diagnosis, resulting in improved patient outcomes and better resource allocation for medical professionals.

B. Linear Regression

Linear regression is a statistical method that is used to predict the relationship between two variables. In the context of breast cancer prediction, linear regression can be used to determine the relationship between risk factors and the likelihood of developing breast cancer. This information can be used to create a predictive model that can help identify individuals who are at high risk for developing breast cancer. The significance of linear regression in breast cancer prediction lies in its ability to analyze complex data sets and identify important predictors. This information can be used to develop a risk assessment tool that can be used by healthcare professionals to identify women who are at high risk for breast cancer. By early detection and diagnosis, the prognosis of the disease can be improved, and better outcomes can be achieved. Linear regression can be used to evaluate the effectiveness of various breast cancer screening and prevention programs. This information can be used to optimize these programs and ensure that resources are being used effectively to reduce the burden of breast cancer.

C. SVM

Support Vector Machine (SVM) algorithm is one of the powerful machine learning techniques used in data science and AI. In the context, the breast cancer prediction, SVM plays a crucial role in detecting early signs of cancer and predicting the risk of developing the disease. SVM algorithms use a boundary line to separate data points into different classes. This boundary limit is optimized to maximize the margin between the classes, which leads to high accuracy in predictions. SVM algorithms can handle non-linear data patterns, which is critical in the case of breast cancer prediction where data is often complex and non-linear. SVM algorithms are capable of handling large datasets, making it ideal for breast cancer prediction where large amounts of data are collected and analyzed. SVM algorithms can effectively select the most relevant features in a dataset, reducing the risk of overfitting and increasing the accuracy of predictions. SVM algorithms can be used for both binary and multi-class classification problems, making it a versatile tool in the field of breast cancer prediction.

Deep learning became one of the increasingly important in the field of breast cancer prediction due to its ability to analyze and process vast amounts of complex and high-dimensional medical data. Deep learning algorithms are capable of identifying complex patterns and relationships in medical data that are not easily noticeable by human experts, resulting in a higher level of accuracy in breast cancer predictions. Deep learning algorithms can process vast amounts of medical data in a matter of seconds, making the process of breast cancer prediction much faster and more efficient. With deep learning, predictions can be made without the need for human input, reducing the possibility of human error and increasing the speed and efficiency of the process. Deep learning can be integrated with other technologies such as computer vision and artificial intelligence to create more comprehensive and accurate predictions. By providing more accurate and faster predictions, deep learning has the potential to improve patient outcomes by enabling early identification of breast cancer and treatment.

TABLE IV: BREAST CANCER CLASSIFICATION USING MACHINE LEARNING

Authors	**Title**	**Journal**	Methods	Results
Amrane *et al.* [3]	"Breast cancer classification using machine learning"	"2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT")"	Naïve Bayesian Classifier (NBC), k-Nearest Neighbors (KNN)	Reviewed and compared the efficiency and accuracy of KNN and NBC.KNN has a lower efficiency of 97.51% and NBC has good accuracy of 96.19%.
Madhu and Vijendra [4]	"Breast Cancer Prediction system"	"Procedia Computer Science"	SVM, Linear Regression and KNN	Suggested an approach that drastically lowers the cost of treatment and raises quality of life by identifying breast cancer at an early stage of development
Mojrian *et al.* [5]	"Hybrid Machine Learning Model of Extreme Learning Machine Radial basis function for Breast Cancer Detection and Diagnosis; a Multilayer Fuzzy Expert System"	"2020 RIVF International Conference on Computing and Communication Technologies (RIVF)"	Extreme Learning Machine (ELM), Support Vector Machine (SVM)	The proposed multilayer fuzzy expert system's reported accuracy for the ELM-RBF model during training, testing, and validation was 99.72%, 99.23%, and 95.69%, respectively.
Bazazeh and Shubair [6]	"Comparative study of machine learning algorithms for breast cancer detection and diagnosis"	"2016 5th International Conference on Electronic Devices, Systems and Applications (ICEDSA)"	Support Vector Machine (SVM), Random Forest (RF), Bayesian Networks (BN)	RFs have the highest possibility of correctly diagnosing tumors, according to the comparison analysis.
Gayathri and Sumathi [7]	"Comparative study of relevance vector machine with various machine learning techniques used for detecting breast cancer"	"2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)"	Relevance Vector Machine (RVM), Support Vector Machine (SVM), Bayesian Networks (BN)	Demonstrates that when compared to other machine learning techniques like SVM, BN, and Fuzzy, the RVM has the highest accuracy.
Islam *et al.* [8]	"Prediction of breast cancer using support vector machine and K-Nearest neighbors"	"2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)"	Support Vector Machine [SVM] and K-Nearest Neighbors [KNN]	This technique demonstrates that SVM produces the best training and testing outcomes. Compared to KNN, this approach has the highest accuracy performance (98.57%).
Amrane, *et al.* [9]	"Breast cancer classification using machine learning"	"2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)"	Naive Bayes (NB) classifier and K-Nearest Neighbors (KNN)	observed that KNN had a higher efficiency of 97.51% but even NB had a good accuracy of 96.19%. However, the running time of the KNN would increase as the dataset grew.
Turgut *et al.* [10]	"Breast cancer outcome prediction with tumour tissue images and machine learning"	"Breast Cancer Res Treat"	Support Vector Machine, K-Nearest Neighbors, Multi-Layer Perceptron, Decision Trees, Random Forests, Logistic Regression, Adaboost and Gradient Boosting Machines	The SVM algorithm, which was employed in the comparison, produces the most accurate results of all the machine learning algorithms.
Gupta and Gupta [11]	"A Comparative Study of Breast Cancer Diagnosis Using Supervised Machine Learning Techniques"	"2018 Second International Conference on Computing Methodologies and Communication (ICCMC)"	Multilayer perceptron [MLP], Decision Tree [DT], Support vector machine [SVM]and K-nearest neighbor [K-NN].	Compared various machine learning techniques for diagnosing BC and discovered MLP performed 98.12% better than other ML strategies.
Alakwaa *et al.* [12]	"Deep Learning Accurately Predicts Estrogen Receptor Status in Breast Cancer Metabolomics Data"	"Journal of Proteome Research"	Deep Learning (DL), Random Forest (RF), Support Vector Machines (SVM), Recursive Partitioning and Regression Trees (RPART), Linear Discriminant Analysis (LDA), Prediction Analysis For Microarrays (PAM), And Generalized Boosted Models (GBM)	Demonstrate that, for categorizing the ER state of breast cancer metabolomics data, DL outperforms other machine learning methods.
Shah and Jivani [14]	"Comparison of data mining classification algorithms for breast cancer prediction"	"Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)"	Decision tree [DT], Bayesian Network [BN], and K-Nearest Neighbor algorithms [KNN]	Proved that Naïve Bayes is the superior algorithm which gives more accuracy and less execution time compared to others.
Ganggayah *et al.* [15]	"Predicting factors for survival of breast cancer patients using machine learning techniques"	"BMC Medical Informatics and Decision Makink"	Decision Tree [DT], Random Forest [RF], Neural Networks [NN], Extreme Boost [EB], Logistic Regression [LR], And Support Vector Machine [SVM]	Presented the analysis of prominent factors of breast cancer for predication of survival rate.
Wang *et al.* [16]	"Identification of Differentially Expressed Genes between Original Breast Cancer and Xenograft Using Machine Learning Algorithms"	"Genes (Basel). 2018"	Monte Carlo feature selection (MCFS), random forest (RF), Support Vector Machine [SVV], and rough set-based rule learning.	Created a useful tool to distinguish natural tumors from PDX tumors based on gene expression levels and use SVM and RF techniques to identify all gene expression changes and forming properties in the murine microenvironment.

Authors	**Title**	**Journal**	Methods	Results
Cain *et al.* [17]	"Multivariate machine learning models for prediction of pathologic response to neoadjuvant therapy in breast cancer using MRI features: a study using an independent validation set"	"Breast Cancer Res Treat"	Logistic Regression [LR] and Support Vector Machine [SVM]	The model made accurate predictions, highlighted the possible application of pretreatment breast, and gave patients access to more individualized care and better quality of life.
Patrício *et al.* [18]	"Using Resistin, glucose, age and BMI to predict the presence of breast cancer"	"BMC Cancer"	Random Forest [RF], Monte Carlo Cross-Validation approach, Logistic Regression [LR], And Support Vector Machine [SVV]	Study proposed a model for breast cancer detection based on biomarkers on a test data set with sensitivity ranging between 82 and 88% and specificity ranging between 85 and 90%.
Singh and Raj [19]	"Breast Cancer Analysis and Prediction by Using Machine Learning"	"International Journal of Research in Engineering and Science"	Random Forest {RF}, Decision Tree[DT], Machine Learning[ML]	The primary goal is to evaluate the accuracy and time required to classify the data correctly, which is used to assess the effectiveness of such algorithms. This outcome of the studies demonstrates that the Random Forest algorithm has the best accuracy rate (99.76%) and the lowest error rate.
Park *et al.* [20]	"Robust predictive model for evaluating breast cancer survivability"	"Engineering Applications of Artificial Intelligence"	Support Vector Machine (SVM), Artificial Neural Networks [ANN], and Semi supervised Learning Models [SSLM]	The importance of a model's stability was discussed in the study, which compared its performance and recommended the optimal model. Overall, the SSLM demonstrated good accuracy and remained stable as model parameters changed. More data is derived from the unlabeled data via SSLM.
Ming *et al.* [21]	"Machine learning-based lifetime breast cancer risk reclassification compared with the BOADICEA model: impact on screening recommendations"	"British journal of cancer"	Markov Chain Monte Carlo [MCMC] , Generalized Linear Mixed Model [GLMM], Adaptive Boosting [ADA] and Random Forest [RF]	A study compared the BOADICEA model's and ML algorithms' breast cancer prediction accuracy. A study found that ML algorithms have better predicted accuracy. The program evaluated family history, cancer pathology, and clinic-demographic data on a dataset of n = 112,587 individuals from 2481 families.
Behravan *et al.* [22]	"Predicting breast cancer risk using interacting genetic and demographic factors and machine learning"	"Scientific Reports"	Linear Regression [LR], and Logistic Regression [LR]	The study demonstrates that demographic risk factors, such as apoptosis, angiogenesis, and oestrogen-related metabolism, are exclusively more relevant than genetic variants when predicting breast cancer risk using machine learning algorithms.
Turkki *et al.* [23]	"Breast cancer outcome prediction with tumour tissue images and machine learning"	"Breast Cancer Res Treat"	Univariate Analysis and Multivariate Analysis	The study discovered that the whole predictive capacity of tumor shape may be used by a machine learning approach to estimate the BC risk using just a huge and complicated image of a tumor sample.
Gaurav Singh [24]	"Breast Cancer Prediction Using Machine Learning"	"International Journal of Scientific Research in Computer Science, Engineering and Information Technology"	Linear Regression [LR], K-nearest neighbor [K-NN], Support vector machine [SVM], and Gaussian Naive Bayes [NB]	The study, which used the Wisconsin breast cancer dataset, discovered that, when compared to SVM, LR, and NB for the provided dataset, nearest neighbor has the highest accuracy for breast cancer prediction.
Chaurasia and Pal [25]	"Applications of Machine Learning Techniques to Predict Diagnostic Breast Cancer"	"SN Computer Science, Springer"	Classification and Regression Tree [CART], Support Vector Machine [SVM], Naïve Bayes [NB], K-Nearest Neighbors [KNN)] Linear Regression [LR] and Multilayer Perceptron [MLP]	The study's major goal was to use statistical analysis to estimate how accurate the feature selection would be. From 32 features, the dataset has reduced to 12 features. To determine the best accuracy, various classifiers were used to reduce characteristics.
Sharma *et al.* [1]	"Various types and management of breast cancer: an overview"	"Journal of advanced pharmaceutical technology & research"	Naïve Bayesian Classifier [NBC], k-Nearest Neighbors [KNN]	The algorithms divide the data for cross validation and testing stages to develop precise and dependable classifiers to determine accuracy and time using the sample of N examples and their classes that is provided.
Sun *et al.* [2]	"A Multimodal Deep Neural Network for Human Breast Cancer Prognosis Prediction by Integrating Multi-Dimensional Data"	"IEEE/ACM Transactions on Computational Biology and Bioinformatics"	Deep Neural Network [DNN]	For the prognostic prediction of breast cancer, the study suggests a Multimodal Deep Neural Network by Combining Multi-dimensional Data (MDNNMD). According to the findings of the thorough performance evaluation, the suggested method outperforms existing methods and prediction algorithms that use one-dimensional data.

IV. Conclusion and Future Improvement

The Current review has given many insights about various machine learning methods / data analytics / algorithms that are used in the prediction of breast cancer or other types of diseases. Most of these methods are popular and have already proven in the mentioned studies with the results that are accepted at various platforms. It was observed that the machine learning models or data analytic algorithms such as Support Vector Machines (SVM), K-Nearest neighbors (KNN), Naïve Bayesian (NB), Bayesian Networks, Decision Trees (DT), Linear Regressions (LR) etc., were used effectively with various papers. Various datasets were effectively used by applying these algorithms to get effective results. There were papers with utilization of multiple algorithms with in one research, which has seen improvement in the accuracy of the results produced.

As a future enhancement to the current study, the current research team is planning to study the Omani Women Health Registry and develop an intelligent risk prediction system that will predict the risk of breast cancer in the local women with various age groups. The system is planned to develop using the algorithms that are hybrid in nature. The hope is get better and accurate results in predicting the breast cancer in Oman Women by providing required parameters.

Acknowledgment

We thank the deanship, management, and research and consultancy committee of UTAS Nizwa, for accepting the research and proposing for funding. We extend our thanks to The Research Council for funding the research. Words are not enough to thank the involvement and support of Mrs. Suad Al Riyami, Mrs. Alice Lontok, and Mrs. Manitha who were the driving force in various ways at various stages of the research. Thanks to the members of the research team for their excellent support.

Funding

This work has been carried out based with the funds allocated by The Research Council, Oman, under the Research Grant funding that is approved with ID: BFP/RGP/ICT/20/142

Conflict of Interest

There are no personal, financial or professional conflicts among authors, or any other entities / groups involved in the research.

References

[1]. Sharma GN, Dave R, Sanadya J, Sharma P, Sharma KK. (2010). Various types and management of breast cancer: an overview. *Journal of advanced pharmaceutical technology & research.* 2010; 1(2): 109–126.

[2]. Sun D, Wang M, Li A. A Multimodal Deep Neural Network for Human Breast Cancer Prognosis Prediction by Integrating Multi-Dimensional Data. In *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, May-June 2019; 16(3): 841-850. doi: 10.1109/TCBB.2018.2806438.

[3]. Amrane M, Oukid S, Gagaoua I, Ensari T. Breast cancer classification using machine learning. *2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)*; 2018: 1-4, doi: 10.1109/EBBT.2018.8391453.

[4]. Madhu K, Vijendra S, Breast Cancer Prediction system. *Procedia Computer Science*, 2018; 132: 371-376. ISSN 1877-0509. https://doi.org/10.1016/j.procs.2018.05.197.

[5]. Mojrian S, Pinter G, Joloudari JH, Felde I, Szabo-Gali A, Nadai L, Mosavi A. Hybrid Machine Learning Model of Extreme Learning Machine Radial basis function for Breast Cancer Detection and Diagnosis; a Multilayer Fuzzy Expert System. In *2020 RIVF International Conference on Computing and Communication Technologies (RIVF),* 2020. https://doi.org/10.1109/rivf48685.2020.9140744.

[6]. Bazazeh D, Shubair R. Comparative study of machine learning algorithms for breast cancer detection and diagnosis. *5th International Conference on Electronic Devices, Systems and Applications (ICEDSA)*, 2016: 1-4, doi: 10.1109/ICEDSA.2016.7818560

[7]. Gayathri BM, Sumathi CP. Comparative study of relevance vector machine with various machine learning techniques used for detecting breast cancer. *IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, 2016: 1-5, doi: 10.1109/ICCIC.2016.7919576.

[8]. Islam MM, Iqbal H, Haque MR, Hasan MK. Prediction of breast cancer using support vector machine and K-Nearest neighbors. *IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*, 2017: 226-229, doi: 10.1109/R10-HTC.2017.8288944.

[9]. Amrane M, Oukid S, Gagaoua I, Ensari T. Breast cancer classification using machine learning. *Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)*, 2018: 1-4, doi: 10.1109/EBBT.2018.8391453.

[10]. Turgut S, Dagtekin M, Ensari T. Microarray breast cancer data classification using machine learning methods. *Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)*, 2018: 1-3, doi: 10.1109/EBBT.2018.8391468.

[11]. Gupta M, Gupta B. A Comparative Study of Breast Cancer Diagnosis Using Supervised Machine Learning Techniques. *Second International Conference on Computing Methodologies and Communication (ICCMC)*, 2018: 997-1002, doi: 10.1109/ICCMC.2018.8487537.

[12]. Alakwaa FM, Chaudhary K, Garmire LX. Deep Learning Accurately Predicts Estrogen Receptor Status in Breast Cancer Metabolomics Data. *J Proteome Res.,* 2018 Jan 5;17(1):337-347. doi: 10.1021/acs.jproteome.7b00595.

[13]. Rehman O, Zhuang H, Muhamed Ali A, Ibrahim A, Li Z. Validation of miRNAs as Breast Cancer Biomarkers with a Machine Learning Approach. *Cancers*, 2019; 11(3): 431. https://doi.org/10.3390/cancers11030431.

[14]. Shah C, Jivani AG. Comparison of data mining classification algorithms for breast cancer prediction. *Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT),* 2013: 1-4, doi: 10.1109/ICCCNT.2013.6726477.

[15]. Ganggayah MD, Taib NA, Har YC, Lió P, Dhillon SK. Predicting factors for survival of breast cancer patients using machine learning techniques. *BMC Medical Informatics and Decision Making*, 2019; 19(1). https://doi.org/10.1186/s12911-019-0801-4.

[16]. Wang D, Li JR, Zhang YH, Chen L, Huang T, Cai YD. Identification of Differentially Expressed Genes between Original Breast Cancer and Xenograft Using Machine Learning Algorithms. *Genes (Basel)*. 2018 Mar 12;9(3):155. doi: 10.3390/genes9030155. PMID: 29534550; PMCID: PMC5867876.

[17]. Cain EH, Saha A, Harowicz MR, Marks JR, Marcom PK, Mazurowski MA. Multivariate machine learning models for prediction of pathologic response to neoadjuvant therapy in breast cancer using MRI features: a study using an independent validation set. *Breast Cancer Research and Treatment*, 2019; 173(2), 455–463. https://doi.org/10.1007/s10549-018-4990-9.

[18]. Patrício M, Pereira JA, Crisóstomo J, Matafome P, Gomes M, Seiça R, Caramelo F. Using Resistin, glucose, age and BMI to predict the presence of breast cancer. *BMC Cancer*, 2018; 18(1). https://doi.org/10.1186/s12885-017-3877-1.

[19]. Singh H, Raj R. Breast Cancer Analysis and Prediction by Using Machine Learning. *International Journal of Research in Engineering and Science (IJRES)*, 2021; 9(6): 69-73.

[20]. Park K, Ali A, Kim D, An Y, Kim M, Shin H. Robust predictive model for evaluating breast cancer survivability. *Engineering Applications of Artificial Intelligence*, 2013; 26(9): 2194–2205. https://doi.org/10.1016/j.engappai.2013.06.013.

[21]. Ming C, Viassolo V, Probst-Hensch N, Dinov ID, Chappuis PO, Katapodi MC. Machine learning-based lifetime breast cancer risk

reclassification compared with the BOADICEA model: impact on screening recommendations. *British Journal of Cancer*, 2020; 123(5): 860–867. https://doi.org/10.1038/s41416-020-0937-0.

[22]. Behravan H, Hartikainen JM, Tengström M, Kosma V, Mannermaa A. Predicting breast cancer risk using interacting genetic and demographic factors and machine learning. *Scientific Reports*, 2020; 10(1). https://doi.org/10.1038/s41598-020-66907-9.

[23]. Turkki R, Byckhov D, Lundin M, Isola J, Nordling S, Kovanen PE, Verrill C, *et al.* Breast cancer outcome prediction with tumour tissue images and machine learning. *Breast Cancer Research and Treatment*, 2019; 177(1): 41–52. https://doi.org/10.1007/s10549-019-05281-1.

[24]. Singh G. Breast Cancer Prediction Using Machine Learning. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT)*, 2020; 6(4):278-284. https://doi.org/10.32628/CSEIT206457.

[25]. Chaurasia V, Pal S. Applications of Machine Learning Techniques to Predict Diagnostic Breast Cancer. *SN Computer Science*, 2020; 1(5). https://doi.org/10.1007/s42979-020-00296-8.

[26]. Duffy SW, Morrish O, Allgood P, Black R, Gillan MGC, Willsher P, Cooke J, *et al.* Mammographic density and breast cancer risk in breast screening assessment cases and women with a family history of breast cancer. *European Journal of Cancer*, 2018; 88: 48–56. https://doi.org/10.1016/j.ejca.2017.10.022.

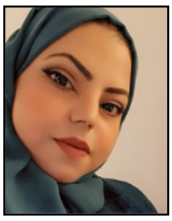
Dr. Tamilselvi Madeswaran holds a Ph.D. degree, from Anna Univeristy. She worked at Lecturer and Senior Lecturer in the Sona College of Technology from 2000 to 2008. She served in the TULEC Computer Education Center as Programmer and acted as Guest Lecturer in the Government College, Salem, India. She is also worked as Academic councellor and University coordinator in the Indira Gandhi Open University and TamilNadu Open University. She is also a member in Computer society of India. She serves as reviewer for well-known national and international journals. She is currently working as Lecturer in IT under Department information technology, University of Technology and Applied Sciences - Nizwa, Sultanate of Oman.
She works in the field of Fuzzy Logic, Semantic Web, machine learning, data analytics and Networks. She Received Appreciation award for the producing Best Results in the University Exam. She has 22 years of teaching experience and 15 years of Research Experience.

Mr. Aruna Kumar Kavuru pursed Bachelors of Computer Science from Nagarjuna University in the year 1997, Masters in Computer Applications in AY 2000 from Madras University and M.Phil in computer science from Periyar univeristy in the year 2013. He is currently pursuing Ph.D. from SCSVMV university, Kanchipuram, India and working as Lecturer in IT under Department information technology, University of Technology and Applied Sciences - Nizwa, Sultanate of Oman. His main research work focuses on Big Data Analytics, Data Mining, IoT and Information Retrieval. He has 22 years of teaching experience and 11 years of Research Experience.

Dr. Padma Theagarajan is a Professor and Head at the Department of Computer Applications, Sona College of Technology, India. Her research interests include machine learning, data analytics and interaction design. She has executed research projects as Principal Investigator funded by AICTE, UGC and NCW. She has published around 60 journal papers, books and book chapters in the fields of knowledge based systems, operations research, data analytic and computational intelligence.

She serves as editor and reviewer for well-known national and international journals. She is very instrumental in setting up the currently associated institution's management information system; developed different software solutions for academics, plantations, dairy forms and textiles on web-based architecture. She is a certified Six Sigma Black Belt. Dr. Padma is a Fellow of the Computer Society of India and Indian Society for Technical Education. She was the recipient of the Shayesta Akhtar Memorial National Award of ISTE in 2015 and uLektz Wall of Fame recognition, awarded to Top 20 Influential Women Educators in Tamilnadu in 2019. Her biography was included in the 2010 edition of the Marquis Who's Who in the World, New Jersey, USA.

Dr. Nasser Al Hadrami Nasser Al-Hadhrami is a Senior Lecturer at the University of Technology and Applied Sciences (UTAS), Oman. Nasser achieved his PhD in Computing Sciences in 2021 from the University of Aberdeen, UK. He obtained his Master's degree in 2014 from the University of Portsmouth (UK) in the field of Cybersecurity. Before joining UTAS, Nasser was working as a Lecturer at the University of Nizwa, Oman, and before this, he had been working as a teacher in the Ministry of Education. Nasser has research interests in the domains of Cyber security and Artificial Intelligence. Mainly, he is interested in the topics of computer and network security, formal methods, and AI in security and security in AI.



Ms. Maya Al Foori is a Specialized Nurse in Oncology Nursing and Clinical Educator, National Oncology Center, Royal Hospital, Oman. She is interested to join research regarding cancer aiming to improve cancer care, and develop screening programs. She is a data collector in a master degree research with a researcher on process.

She is active member in cancer awareness activities arranged by NOC and per invitation. She designed patient educational materials about cancer and cancer treatment side effect management as well awareness pamphlets and posters for NOC patient care. Her job focused on training nurses in oncology nursing care and conducting other related courses.

Mr. Ohm Rambabu pursed Bachelor of Engineering in Electronics and Communications first Class from Bharthidasan University, Completed in 1999, Master of Technology in Electronics and Communication First Class from Manav Bharthi University completed in 2012 He is currently working as Technical Support Specialist, under ETC department, University of Technology and Applied Sciences - Nizwa, Sultanate of Oman. He is Cisco Certified Network Associate Routing and Switching technician. He has 21 years of and expertise in Operations management, Supervision, Training and Development, Workforce management, Quality management as well as enhance department efficiencies amidst fast- paced working environments.