Machine Learning Techniques for Breast Cancer Analysis: A Systematic Literature Review

Lina Alkhathlan[1] and Abdul Khader Jilani Saudagar[2]

Information Systems Department, College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia

Summary
Breast cancer (BC) is one of the most common cancers and is known to be the leading cause of death among females around the world. Breast cancer occurs when cells in the breast develop a malignant tumor. Detecting BC at an early stage with state-of-the-art technologies helps in treating BC and reduces the risk of death. Currently, mammography is the most commonly used technique for detecting BC. In order to improve the mammogram analysis, researchers have studied the feasibility of using artificial intelligence to help doctors in detecting any changes that may lead to cancer. This review paper investigates utilizing machine learning (ML) algorithms for BC prediction and classification, which will be beneficial in early diagnosis and treatment for BC and for future researchers in exploring the different ML techniques and selecting the most suitable one for their future research.

Key words:
Artificial neural network, Breast cancer, Machine learning, Support vector machine.

1. Introduction

One of the most valuable aspects of a fulfilling life is good health. Maintaining a healthy body, not only physically but also mentally and psychologically, is essential. An unhealthy body is easily attacked by various diseases that could be fatal, cause discomfort, or cause pain and suffering that last for many years. Each disease has certain symptoms and signs and can be either infectious or noninfectious. One serious noninfectious disease is cancer. According to [51], cancer occurs when body cells divide without control and can reach nearby tissues. If this division occurs in the breast cells, a malignant tumor can develop, which is called breast cancer (BC). According to the World Health Organization, BC is one of the most common cancers and the leading cause of death among females around the world [7] [45]. It is also the most frequent cancer in developed and less-developed countries. Diseases are major burden for all societies. According to [42], the burden of disease rates across the world varied from 40,000 to 70,000 disability-adjusted life years (DALYs) for every 100,000 individuals in high-burden countries in 2016. The burden of disease rate for North Africa and the Middle East was 31,321.32 DALYs per 100,000 individuals; the rate reached 25,296.55 DALYs per 100,000 individuals in Saudi Arabia in 2017. The incidence of BC ranged from 19.3 for every 100,000 females in Eastern Africa to 89.7 for every 100,000 females in Western Europe [5]. For developed countries excluding Japan, the incidence rates are high (more than 80 per 100,000), while the rates are lower than 40 per 100,000 in most less-developed countries [53].

In the Arab world, based on statistics shown in [19], the incidence of BC increased over the past 26 years and is expected to keep increasing if no actions are taken. In 2016, the number of new BC cases reached 45,980 (28/100,000), with 20,063 (11/100,000) deaths in the region. Previous studies have shown that BC presents at an earlier age and at an advanced stage in Arab females [19]. In Saudi Arabia, statistics provided by the Saudi Cancer Registry of the King Faisal Specialist Hospital and Research Centre showed that 930 new cases of BC are diagnosed each year [3].

According to [20], most BC cases are detected at an advanced stage, when the disease has progressed and the cure rate is reduced. The cure rate is more than 90% if the disease is detected early. The lack of health awareness increases BC incidence in Saudi Arabia. It has also been observed that BC mortality rates are lower in developed countries due to the existence of early detection programs. Detecting and treating BC early helps to reduce the risk of death due to the disease. Therefore, periodic health checkups are essential, as they minimize the time between diagnosis and treatment. However, periodic health checkups are sometimes insufficient for early diagnosis of BC, as changes in these cases cannot be detected by the unaided eye or by blood tests. Accordingly, more sophisticated methods are needed to detect BC in the earlier stages. Currently, mammography, known as the gold standard in imaging, is the most frequently used technique for detecting BC early. In order to improve mammogram analysis, many researchers have studied the feasibility of using artificial intelligence (AI) to help doctors in detecting any harmful changes that may lead to cancer. The majority of these studies focused on using machine learning (ML) techniques, which create

Manuscript received June 6, 2020
Manuscript revised June 20, 2020

algorithms to detect any changes in the patient and predict the probability of cancer.

One study [14] explained that ML uses algorithms that simulate human intelligence by learning from the environment, that is, the algorithm inputs data to fulfill a required task without explicit programming in order to deliver a specific outcome. ML is classified into three types: supervised, unsupervised, and semi-supervised learning. For supervised learning, data are labeled, and the input data can predict the output data, while data in unsupervised learning are not labeled. Most of the data are not labeled in semi-supervised learning, and supervised and unsupervised techniques are used together. ML has been utilized in various significant applications, such as data mining, image recognition, and expert systems [33].

The focus of this review paper is on diagnosing BC, as the severity of this disease increases the need to utilize ML techniques to diagnose BC. In the next section, the authors present a review of studies of the use of these techniques for the prediction and diagnosis of BC. This review paper contributes to this effort by showing that ML techniques, including artificial neural networks (ANN), decision trees (DT), K-nearest neighbor (KNN) algorithms, support vector machines (SVM), and naïve Bayesian classification, can be effectively used in BC prediction and classification. Furthermore, this paper provides a basis for comparing these techniques that can be helpful for developers in deciding which ML algorithm is suitable for building a specific model. Section 2 of this paper explains the literature review, which includes the research framework and method for collecting articles for conducting the research. Section 3 analyzes the number of publications over a time period which helps the future researchers to get familiar which latest developments in this domain of research and their venues (conference, journal, etc.). This is followed by a classification of articles based on three areas: ML algorithms, data sources, and the performance accuracy of ML algorithms. Conclusions are presented in Section 4.

2. Literature Review

2.1 Research framework

The authors performed a systematic literature review to survey state-of-the-art ML techniques used for prediction and classification processes. This research answers the question,

How can ML help in predicting and classifying BC?

The literature review was performed using Google scholar search, the Saudi Digital Library, and multiple English-language databases, such as Springer, Institute of Electrical and Electronics Engineers (IEEE) Xplore, Science Direct, Elsevier, etc. The authors considered papers published between 2009 and 2018, using the keywords machine learning, breast cancer, breast cancer prediction, artificial neural network, support vector machine, decision trees, naïve Bayesian, and K-nearest neighbor. After excluding insignificant studies, the authors reviewed 32 papers.

2.2 Survey of number of publications and venue

The authors of this research selected papers that were published at conferences and in journals, proceedings, and books and excluded patent applications, lecture notes, slides, and information websites. Fig. 1 lists the selected papers on predicting and classifying BC using ML with their year of publication.

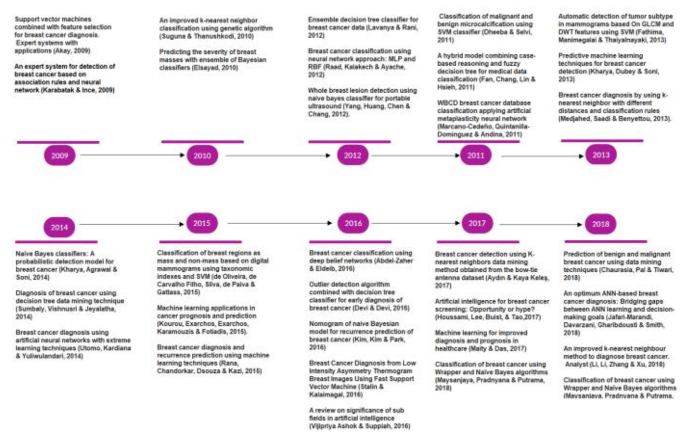

Fig. 1 Published articles in predicting and classifying breast cancer using machine learning with the year of its publication

3. Classification of Articles

3.1 Machine learning algorithms

BC detection is commonly performed with the use of ML algorithms. According to [26], frequently used ML algorithms include ANN, DT, KNN, SVM, and naïve Bayesian. The next subsections present the ML algorithms used by researchers in the selected articles on predicting BC.

3.1.1 Artificial neural networks

Artificial neural networks (ANN) are interesting and powerful ML-based techniques used in various fields to process, analyze, and predict data [14] [16]. An ANN simulates the human brain to learn complex tasks. It is presented as an interconnected set of nodes that looks like the network of neurons in the human brain [16]. In 2009, researchers in [24] introduced an expert system that detects BC using association rules (AR) and ANN. The study used AR to minimize the database dimensions and used ANN for classification. The classification rate of the proposed AR with ANN is compared to the ANN model. This study used the Wisconsin Breast Cancer Database, which contains 9 attributes and around 699 records.

In 2011, [34] proposed a metaplasticity ANN that was inspired by information theory developed by Shannon. In their study, the artificial metaplasticity multilayer perceptron (AMMLP) algorithm is proposed to classify BC as benign or malignant. To update the weights, the AMMLP algorithm gives updating priority to activations with lower frequency than activations with higher frequency, resulting in higher efficiency in the training phase. Using the AMMLP algorithm, the archived classification accuracy was 99.26%. MAT-LAB software was used to implement the proposed AMMLP as a classifier. A year later, authors in [39] proposed an approach to BC classification as benign or malignant using neural networks, particularly the Multilayer Perceptron and Radial Basis Function (RBF). When the proposed approach was tested on the Wisconsin Breast Cancer Database, the classification accuracy of the RBF neural network was 97%. In 2014, research [48] used extreme learning techniques in implementing ANN for diagnosing BC. A comparison of performance was conducted between extreme ML artificial neural networks and back-propagation artificial neural networks, in which specificity, sensitivity, and accuracy were considered. Researchers concluded that extreme learning machine artificial neural networks in general provided better results in diagnosing BC than the back-propagation artificial neural networks.

A study [1] in 2016 proposed a deep belief network (DBN) combined with ANN (DBN-ANN) to improve the BC detection accuracy of a computer-aided diagnosis system. The weights were initialized from the DBN path. The architecture used to test the database consisted of nine inputs, with four hidden and then two hidden, providing one output. The accuracy of using a scaled conjugate gradient reached 99.59% with training, and data validation (DBN-ANN) reached 63.84–36.16%, where the accuracy percentage of applying Randomly Initialized Weight Back-Propagation Neural Network (RIW-BPNN) using Levenberg-Marquardt case reached 99.68% with training, and validating data reached 61.35–38.65%. MAT-LAB 2014a software was used for implementation. In 2018, [23] proposed a Life-Sensitive Self-Organizing Error Driven (LS-SOED) classification method that enhances the ANN decision-making process. Researchers believe that higher accuracy does not always provide the best decision in diagnosis. LS-SOED combines the use of self-organizing map, multilayer perceptron, and fluid genetic algorithms. The study results show that the diagnosis decisions made using LS-SOED saved 30 years and 8 years of life from datasets which has 283 and 57 patients respectively. The implementing tool for the proposed method is MAT-LAB software.

3.1.2 Support vector machine

The second algorithm to be discussed is the support vector machine (SVM), which finds an (N – 1) dimensional hyperplane from a given set of points of two types in an N-dimensional place to differentiate the two classes [44]. In 2009, researchers in [2] proposed an SVM algorithm including feature selection to diagnose BC. The accuracy of classification reached 99.51% using an SVM model that contained five features. In 2011, researchers at [13] presented a new classification approach to detect microcalcification clusters in digital mammograms with the use of SVM. The results of the study indicated that the proposed approach's accuracy was 86.1%, which is greater than the accuracy of other classification approaches reviewed in the paper.

In 2013, researchers at [18] proposed a methodology that classifies tumors as benign or malignant. An SVM algorithm was trained using the MATLAB bioinformatics toolbox. The obtained classification accuracy rate reached 95%. Two years later, researchers at [11] proposed a methodology that discriminates and classifies regions extracted from mammograms as mass and non-mass using SVM. Their results showed that the accuracy of successfully classifying the masses and non-masses was 98.88%. The tool used was LibSVM software, which is a library for SVMs. In 2017, a study [37] proposed an automatic diagnostic method by combining a two-step clustering algorithm and SVM to detect the hidden patterns of malignant and benign tumors with 99.1% classification accuracy. Researchers proved that this approach improves the prediction accuracy rate and minimizes misclassification error in cancer. A year later, [50] used an ensemble method that included two types of SVM and six types of kernel functions for BC diagnosis. The accuracy of the proposed model reached 97.10% using the Wisconsin Diagnostic Breast Cancer dataset. R packages for data science were used to implement SVM models, while the ensemble model was implemented using R programming language.

3.1.3 Decision trees

The third algorithm is the DT, which is defined by [44] as a predictive model that uses a tree-like graph to predict the outcome of a variable by applying decision rules derived from the features of the data. In 2011, researchers in [17] developed a hybrid model that encompassed two methods: a fuzzy decision tree for classification and a case-based reasoning method called case-based reasoning and fuzzy decision tree (CBFDT) to classify medical records. The average accuracy of predicting BC using the CBFDT model is up to 98.4%. The proposed model can provide accurate but also comprehensible decision rules that help physicians to reach effective conclusions in medical diagnosis. Researchers applied SPSS statistical software for stepwise regression analysis. A year later, researchers at [30] employed a DT algorithm for classifying medical data and evaluated the accuracy of classifying BC datasets using deferent techniques. The study used WEKA as an implementation tool. A DT-based data mining technique for early detection and classification of BC as benign and malignant was presented in [47]. To establish the model, the J48 DT algorithm is used, which correctly classified 94.5637% of the instances. The J48 algorithm was implemented using the WEKA tool.

In 2016, researchers at [12] proposed an approach with three steps: clustering, pre-processing, and classification. The algorithm was implemented using WEKA, and the classification accuracy of the algorithm reached 99.9%. Two years later, [9] developed models for predicting BC survivability using a naïve Bayes algorithm, RBF neural network, and J48 decision tree. The results indicated that the prediction accuracy of J48 reached 93.41%. WEKA was used for dataset analysis and performance evaluation.

3.1.4 K-nearest neighbor

The fourth algorithm to be discussed was KNN, which classifies data to determine which group a data point belongs to by exploring the data points around it [44]. In 2010, [46] proposed a classifier called genetic KNN (GKNN), which combines the genetic algorithm with the KNN algorithm. Researchers tested the performance of the GKNN classifier with a BC dataset and other medical datasets and concluded that the proposed method minimized the complexity of KNN and improved the accuracy of classification. In 2013, researchers published a paper [36] about a study using a KNN algorithm for classifying BC as benign or malignant with multiple types of rules, such as random and consensus. Study results showed the accuracy of using the KNN algorithm with Euclidean distance reached 98.70% and with Manhattan distance, 98.48%. Two years later, [41] introduced an SVM Sequential minimal optimization (SMO) model implementing the KNN technique with Euclidean distance, Manhattan measures, and other ML techniques to classify whether BC is benign or malignant and to predict the recurrence of a malignant tumor after a certain period of time. The tool used for implementation was MATLAB software using the UCI ML repository. For the overall methodology, results showed that the KNN technique resulted in the greatest accuracy.

In 2017, a research article [4] used the bowtie antenna dataset in the detection of BC using the KNN algorithm. The accuracy rate of using the KNN algorithm reached 90%. The WEKA tool was used to implement the algorithm. First, the data were converted to Attribute-Relation File Format (ARFF), and then an instance-based learning algorithm was used to implement the KNN algorithm. The 10-fold cross was used for validation to achieve the most accurate results using a Java software tool known as Knowledge Extraction based on Evolutionary Learning (KEEL). In 2018, [32] proposed an entropy-weighted local-hyperplane KNN (EWHK) algorithm, which introduced the use of the information entropy of the samples as a classifier to enhance the accuracy of detecting BC. EWHK showed a significant improvement compared to adaptive weighted k-local hyperplane (AWKH) and KNN. The average diagnostic accuracy of the EWHK classifier reached 92.33%. In the same year, [29] designed a BC prediction system by analyzing mini-attributes chosen from the dataset. Researchers concluded that the KNN classifier achieved high accuracy at 99.28%.

3.1.5 Naïve Bayesian

The last algorithm to be discussed is the naïve Bayesian classifier. According to [44], naïve Bayesian applies Bayes theorem and the naive assumption of conditional independence among features. In 2010, the author at [15] used two Bayesian classifiers to predict the severity of breast masses. The accuracy of prediction using the Bayesian algorithm reached 91.83% applied on the training dataset. The ensemble stream was implemented using SPSS software. In 2012, researchers at [52] proposed a system to automatically detect the entire breast lump based on naive Bayes. The lumps identified in the ultrasounds were classified as malignant or normal lumps. Based on the experiment's results in 31 cases with 33 lesions, the accuracy of the system reached 93.4%. The system was implemented using the Microsoft Visual Studio tool and the C++ language. An intelligent BC prediction system based on a naive Bayesian classifier was proposed in 2014 with an accuracy rate that reached 93% [25]. The proposed system was implemented using the Java tool.

Two years later, in [27] a nomogram using naïve Bayesian was developed for predicting recurrent BC within 5 years after treatment and showed 80% performance accuracy. The nomogram can aid physicians and patients for decision-making. The implementation tool used was Orange Toolbox, which is used for data visualization, ML, and data mining. In 2018, researchers at [35] developed a CAD-based scheme by combining two algorithms. Feature selection was performed using a wrapper algorithm, while

BC classification was performed using a naive Bayes algorithm. Using a wrapper method for feature selection was beneficial in identifying the relevant features, which helps in improving the accuracy of classification. The accuracy rate was up to 99.27% where the WEKA framework was used for the implementation.

3.2 Implementation tools

Several tools are used in the implementation of ML algorithms. In [34], [41], [23], and [1], MATLAB software was the tool used to implement the proposed classifier. Researchers at [18] implemented an SVM classifier using MATLAB software. On the other hand, researchers in [11] implement their proposed algorithm using LibSVM software, which is a library for SVMs. Moreover, [50] implemented an SVM model using R packages, while the ensemble structure was coded in the R language. Researchers at [30], [47], [12], [9], [4], and [35] used the Waikato Environment for Knowledge Analysis (WEKA) tool. WEKA contains ML algorithms that perform massive data mining tasks. Researchers in [52] implemented their proposed system with the Microsoft Visual Studio tool using the C++ language. For [25], the proposed system was implemented using Java. Finally, researchers in [27] implemented the naive Bayes classifier and nomogram visualization using Orange Toolbox for data visualization, ML, and data mining. Fig. 2 shows a simple graph of the tools frequently used in related research.

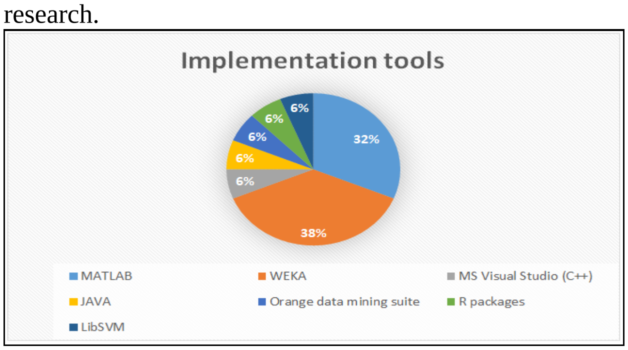

Fig. 2 Implementation tools used for breast cancer diagnosis and classification

3.3 Data sources

This section discusses the size and sources of the data being used by ML algorithms. One of the most frequently used data sources is the Wisconsin Breast Cancer Database, which consists of 699 patients' records gathered by William H. Wolberg in 1991. Each record in the database has nine attributes rated from 1 to 10, where 1 is the most normal state and 10 is the most abnormal state. Each sample is associated with a class label of benign or malignant.

Another data source used in [13] and [18] is the Mammographic Image Analysis Society (MIAS) database. The MIAS database, generated by UK research groups, consists of 322 digitized mammographic images for BC research. The database contains images for 161 patients with breast tumors that may be normal, benign, or malignant. The images are 1024 × 1024 pixels in size, which includes ground truth information on the sites of the abnormal cells shown in the mammogram. Researchers in [11] used a different data source, known as the Digital Database for Screening Mammography (DDSM). DDSM contains digitized mammographic images that are available online. The database consists of 2,620 patient exams, and each exam contains two images of each breast which are marked by specialist physicians and stored in files. Researchers in [50] used two data sources, the Wisconsin Breast Cancer Database and the Surveillance, Epidemiology, and End Results (SEER) BC dataset. The dataset, containing 800,000 instances, is publicly accessible by accepting and signing the required documents.

In [17], researchers used two medical datasets as data resources: liver disorders, Wisconsin Breast Cancer Database are chosen from the UCI database. The liver disorders database consists of 345 observations that represent the blood test result of 345 patients. Researchers at [52] analyzed their results based on real-time data collected from 31 patients. The obtained two-dimensional ultrasound images of each patient were seen by a professional radiologist. The study in [27] used data taken from Ajou University Medical Center in Korea. The dataset consisted of 679 BC patients who had surgery followed by consistent follow-up for at least 5 years. Fig. 3 below is a simple graph of the data sources used frequently in published articles.

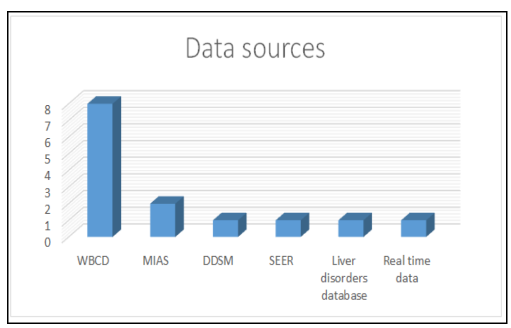

Fig. 3 Data sources used in breast cancer diagnosis and classification

3.4 Machine learning algorithms' performance accuracy

This section aims to demonstrate the comparative performance accuracy of ML algorithms used in related studies. Researchers have used various evaluation methods to assess the performance accuracy of ML algorithms. After reviewing the articles, the authors analyzed that the performance accuracy using ANN in classifying BC ranges from 97% up to 99.68%. The performance accuracy of using an SVM algorithm in classifying BC varied from 86.1% to 99.51%. For the DT technique, the performance accuracy in classifying BC is between 93.41% and 99.9%. For classifying BC using KNN, the performance accuracy rates varied between 90% and 98.70%. Using naïve Bayesian as a classifier reached performance accuracy from 80% up to 99.27%. Fig. 4 presents a graph showing the performance accuracy achieved by ML algorithms in the published articles.

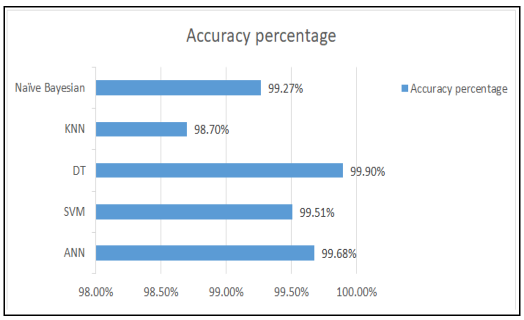

Fig. 4 Performance accuracy of machine learning algorithms

4 Discussion and Results

In this review paper, the authors examined the most recent studies relevant to cancer prediction and diagnosis using ML techniques, and a brief description of ML and some of the classification algorithms currently being used have been presented. Despite the large number of ML studies previously published with accurate results in BC prediction, it is critical to identify the potential drawbacks, such as determining the appropriate data samples to be collected and validating the classification results.

One limitation found in some of the studies reviewed is the small size of the data samples, as samples should be large enough to enable sufficient partitioning of the data into training and testing sets to result in a proper validation.

After reviewing the studies in this paper, we note that SVM is the most common ML algorithm applied in BC prediction due to its accurate performance. Selecting the most appropriate ML algorithm depends on multiple factors, such as data type, data size, time boundaries, and the type of prediction generated [28]. More public databases containing valid cancer datasets are needed to help researchers to conduct studies that result in further improvement in accuracy of the results.

5 Conclusions

In this review paper, the researchers' goal was to find a simpler approach to early diagnosis and classification of BC. The paper discusses ML concepts and summarizes the various ML techniques applied in BC prediction and diagnosis. Based on our analysis of the reviewed papers, no tool currently available diagnoses BC and classifies its stage. Such a tool would be helpful to both patients and doctors, as BC prediction at an early stage increases the chances of treatment for patients and helps doctors to make better decisions in a timely manner. The tool could also be used by medical students to give them a deeper knowledge about the disease.
The application of ML algorithms for pattern classification could be a promising tool for early diagnosis in the BC domain. The most important contribution of this review paper is to provide a comprehensive analysis of recent research on ML to help researchers develop detailed insights into the most recent developments in the field, potential gaps, effective techniques used, and choosing future directions. For future research, the researchers plan to develop an algorithm to predict BC as well as classify the stage of the disease.

References

[1] Abdel-Zaher, A. M., & Eldeib, A. M. (2016). Breast cancer classification using deep belief networks. Expert Systems with Applications, 46, 139-144.

[2] Akay, M. F. (2009). Support vector machines combined with feature selection for breast cancer diagnosis. Expert systems with applications, 36(2), 3240-3247.

[3] Alotaibi, R. M., Rezk, H. R., Juliana, C. I., & Guure, C. (2018). Breast cancer mortality in Saudi Arabia: Modelling observed and unobserved factors. PloS one, 13(10), e0206148.

[4] Aydın, E. A., & Kaya Keleş, M. (2017). Breast cancer detection using K-nearest neighbors data mining method obtained from the bow-tie antenna dataset. International Journal of RF and Microwave Computer-Aided Engineering, 27(6), e21098.

[5] Becker, N. G. (2015). Modeling to inform infectious disease control. Chapman and Hall/CRC.

[6] Bundy, A. (2017). Preparing for the future of Artificial Intelligence.

[7] Cancer country profiles 2014. (2014, December 3). Retrieved October 5, 2019, from https://www.who.int/cancer/country-profiles/en/.

[8] Cancer Facts & Figures 2016. (2016). Retrieved October 5, 2019, from https://www.cancer.org/research/cancer-facts-

statistics/all-cancer-facts-figures/cancer-facts-figures-2016.html.
[9] Chaurasia, V., Pal, S., & Tiwari, B. B. (2018). Prediction of benign and malignant breast cancer using data mining techniques. Journal of Algorithms & Computational Technology, 12(2), 119-126.
[10] Cruz, J. A., & Wishart, D. S. (2006). Applications of machine learning in cancer prediction and prognosis. Cancer informatics, 2, 117693510600200030.
[11] de Oliveira, F. S. S., de Carvalho Filho, A. O., Silva, A. C., de Paiva, A. C., & Gattass, M. (2015). Classification of breast regions as mass and non-mass based on digital mammograms using taxonomic indexes and SVM. Computers in biology and medicine, 57, 42-53.
[12] Devi, R. D. H., & Devi, M. I. (2016). Outlier detection algorithm combined with decision tree classifier for early diagnosis of breast cancer. Int J Adv Engg Tech/Vol. VII/Issue II/April-June, 93, 98.
[13] Dheeba, J., & Selvi, S. T. (2011, March). Classification of malignant and benign microcalcification using SVM classifier. In 2011 International Conference on Emerging Trends in Electrical and Computer Technology (pp. 686-690). IEEE.
[14] El Naqa, I., & Murphy, M. J. (2015). What is machine learning?. In Machine Learning in Radiation Oncology (pp. 3-11). Springer, Cham.
[15] Elsayad, A. M. (2010). Predicting the severity of breast masses with ensemble of Bayesian classifiers. Journal of Computer Science, 6(5), 576.
[16] El-Shahat, A. (2018). Introductory Chapter: Artificial Neural Networks. In Advanced Applications for Artificial Neural Networks. IntechOpen.
[17] Fan, C. Y., Chang, P. C., Lin, J. J., & Hsieh, J. C. (2011). A hybrid model combining case-based reasoning and fuzzy decision tree for medical data classification. Applied Soft Computing, 11(1), 632-644.
[18] Fathima, M. M., Manimegalai, D., & Thaiyalnayaki, S. (2013, February). Automatic detection of tumor subtype in mammograms based On GLCM and DWT features using SVM. In 2013 International Conference on Information Communication and Embedded Systems (ICICES) (pp. 809-813). IEEE.
[19] Hashim, M. J., Al-Shamsi, F. A., Al-Marzooqi, N. A., Al-Qasemi, S. S., Mokdad, A. H., & Khan, G. (2018). Burden of Breast Cancer in the Arab World: Findings from Global Burden of Disease, 2016. Journal of epidemiology and global health, 8(1), 54-58.
[20] Health Days 2018: Breast Cancer Awareness Month. (2018, September 27). Retrieved October 5, 2019, from https://www.moh.gov.sa/en/HealthAwareness/HealthDay/2018/Pages/HealthDay-2018-10-01-31.aspx.
[21] Houssami, N., Kirkpatrick-Jones, G., Noguchi, N., & Lee, C. I. (2019). Artificial Intelligence (AI) for the early detection of breast cancer: a scoping review to assess AI's potential in breast screening practice. Expert review of medical devices, 16(5), 351-362.
[22] Houssami, N., Lee, C. I., Buist, D. S., & Tao, D. (2017). Artificial intelligence for breast cancer screening: Opportunity or hype?. The Breast, 36, 31-33.
[23] Jafari-Marandi, R., Davarzani, S., Gharibdousti, M. S., & Smith, B. K. (2018). An optimum ANN-based breast cancer diagnosis: Bridging gaps between ANN learning and decision-making goals. Applied Soft Computing, 72, 108-120.
[24] Karabatak, M., & Ince, M. C. (2009). An expert system for detection of breast cancer based on association rules and neural network. Expert systems with Applications, 36(2), 3465-3469.
[25] Kharya, S., Agrawal, S., & Soni, S. (2014). Naive Bayes classifiers: A probabilistic detection model for breast cancer. International Journal of Computer Applications, 92(10), 0975-8887.
[26] Kharya, S., Dubey, D., & Soni, S. (2013). Predictive machine learning techniques for breast cancer detection. International journal of computer science and information Technologies, 4(6), 1023-1028.
[27] Kim, W., Kim, K. S., & Park, R. W. (2016). Nomogram of naive Bayesian model for recurrence prediction of breast cancer. Healthcare informatics research, 22(2), 89-94.
[28] 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.
[29] Kumari, M., & Singh, V. (2018). Breast Cancer Prediction system. Procedia computer science, 132, 371-376.
[30] Lavanya, D., & Rani, K. U. (2012). Ensemble decision tree classifier for breast cancer data. International Journal of Information Technology Convergence and Services, 2(1), 17.
[31] Li, D., & Du, Y. (2017). Artificial intelligence with uncertainty. CRC press.
[32] Li, Q., Li, W., Zhang, J., & Xu, Z. (2018). An improved k-nearest neighbour method to diagnose breast cancer. Analyst, 143(12), 2807-2811.
[33] Maity, N. G., & Das, S. (2017, March). Machine learning for improved diagnosis and prognosis in healthcare. In 2017 IEEE Aerospace Conference (pp. 1-9). IEEE.
[34] Marcano-Cedeño, A., Quintanilla-Domínguez, J., & Andina, D. (2011). WBCD breast cancer database classification applying artificial metaplasticity neural network. Expert Systems with Applications, 38(8), 9573-9579.
[35] Maysanjaya, I. M. D., Pradnyana, I. M. A., & Putrama, I. M. (2018). Classification of breast cancer using Wrapper and Naïve Bayes algorithms. In Journal of Physics: Conference Series (Vol. 1040, No. 1, p. 012017). IOP Publishing.
[36] Medjahed, S. A., Saadi, T. A., & Benyettou, A. (2013). Breast cancer diagnosis by using k-nearest neighbor with different distances and classification rules. International Journal of Computer Applications, 62(1).
[37] Osman, A. H. (2017). An enhanced breast cancer diagnosis scheme based on two-step-SVM technique. Int. J. Adv. Comput. Sci. Appl, 8(4), 158-165.
[38] Pertusa, J. F., & Morante-Redolat, J. M. (2019). Main Steps in Image Processing and Quantification: The Analysis Workflow. In Computer Optimized Microscopy (pp. 3-21). Humana, New York, NY.

[39] Raad, A., Kalakech, A., & Ayache, M. (2012). Breast cancer classification using neural network approach: MLP and RBF. networks, 7(8), 9.
[40] Rajappa, S., & Raj, G. (2016, January). Application and scope analysis of Augmented Reality in marketing using image processing technique. In 2016 6th International Conference-Cloud System and Big Data Engineering (Confluence) (pp. 435-440). IEEE.
[41] Rana, M., Chandorkar, P., Dsouza, A., & Kazi, N. (2015). Breast cancer diagnosis and recurrence prediction using machine learning techniques. IJRET: International Journal of Research in Engineering and Technology eISSN, 2319-1163.
[42] Roser, M., & Ritchie, H. (2016, January 25). Burden of Disease. Retrieved October 5, 2019, from https://ourworldindata.org/burden-of-disease.
[43] Russell, S. J., & Norvig, P. (2016). Artificial intelligence: a modern approach. Malaysia; Pearson Education Limited,.
[44] Sakr, S., Elshawi, R., Ahmed, A. M., Qureshi, W. T., Brawner, C. A., Keteyian, S. J., & Al-Mallah, M. H. (2017). Comparison of machine learning techniques to predict all-cause mortality using fitness data: the Henry ford exercIse testing (FIT) project. BMC medical informatics and decision making, 17(1), 174.
[45] Stalin, M. S., & Kalaimagal, R. (2016). Breast Cancer Diagnosis from Low Intensity Asymmetry Thermogram Breast Images Using Fast Support Vector Machine. i-manager's Journal on Image Processing, 3(3), 17.
[46] Suguna, N., & Thanushkodi, K. (2010). An improved k-nearest neighbor classification using genetic algorithm. International Journal of Computer Science Issues, 7(2), 18-21.
[47] Sumbaly, R., Vishnusri, N., & Jeyalatha, S. (2014). Diagnosis of breast cancer using decision tree data mining technique. International Journal of Computer Applications, 98(10).
[48] Utomo, C. P., Kardiana, A., & Yuliwulandari, R. (2014). Breast cancer diagnosis using artificial neural networks with extreme learning techniques. International Journal of Advanced Research in Artificial Intelligence, 3(7), 10-14.
[49] Vijipriya, J., Ashok, J., & Suppiah, S. (2016).A review on significance of sub fields in artificial intelligence. INTERNATIONAL JOURNAL OF LATEST TRENDS IN ENGINEERING AND TECHNOLOGY, 6(3), 542-548.
[50] Wang, H., Zheng, B., Yoon, S. W., & Ko, H. S. (2018). A support vector machine-based ensemble algorithm for breast cancer diagnosis. European Journal of Operational Research, 267(2), 687-699.
[51] What Is Cancer? (2015, February 9). Retrieved October 5, 2019, from http://www.cancer.gov/cancertopics/cancerlibrary/what-is-cancer.
[52] Yang, M. C., Huang, C. S., Chen, J. H., & Chang, R. F. (2012). Whole breast lesion detection using naive bayes classifier for portable ultrasound. Ultrasound in medicine & biology, 38(11), 1870-1880.
[53] Zaidi, Z., & Dib, H. A. (2019). The worldwide female breast cancer incidence and survival, 2018.

Lina Alkhathlan is a Master student currently pursuing MSc in Information Systems from Information Systems Department, College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University.

Abdul Khader Jilani Saudagar received his Bachelor of Engineering B.E, Master of Technology M.Tech and Doctor of Philosophy PhD in Computer Science & Engineering in 2001, 2006 and 2010 respectively. His areas of interests are: Artificial Image Processing, Information Technology, Databases, Web and Mobile Application Development. He has 10 years of research and teaching experience at both undergraduate (UG) and postgraduate (PG) level. Presently working as Associate Professor in Information Systems Department, College of Computer & Information Sciences (CCIS), Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Kingdom of Saudi Arabia. He is also the Head of Intelligent Interactive Systems Research Group (IISRG) at CCIS. He was the principal investigator of the funded projects from KACST, Deanship of Scientific Research (IMSIU) and Research Development Office (RDO) Ministry of Education, Kingdom of Saudi Arabia. Dr. Saudagar has published a number of research papers in International Journals and Conferences. He is associated as member with various professional bodies like ACM, IACSIT, IAENG, ISTE etc., and working as Editorial Board member, Reviewer for many international Journals.