PalArch's Journal of Archaeology of Egypt / Egyptology

BREAST CANCER DIAGNOSIS BASED ON K-NEAREST NEIGHBORS: A REVIEW

Shler Farhad Khorshid[1], Adnan Mohsin Abdulazeez[2]

[1,2]Duhok Polytechnic University, Duhok, Kurdistan Region, Iraq

E-mail: shler.sulayvani@gmail.com, adnan.mohsin@dpu.edu.krd

Shler Farhad Khorshid, Adnan Mohsin Abdulazeez. Breast Cancer Diagnosis Based On K-Nearest Neighbors: A Review-- Palarch's Journal Of Archaeology Of Egypt/Egyptology 18(4), 1927-1951. ISSN 1567-214x

Keyword: Machine Learning, Breast Cancer, Mammogram, K-Nearest Neighbors.

ABSTRACT

The techniques of machine learning are commonly used in classifying breast lesions, as they can improve the mammogram accuracy in detecting malignant masses. One of the top causes of death for women remains breast cancer. Early diagnosis can facilitate adequate treatment and reduce morbidity and mortality. Screening for cancer via mammogram can be an efficient method to detect breast lumps at an earlier stage. The main difficulty occurs during cancer detection and the distinction between a diagnosis to check whether a patient has a malignant or benign type of disease. Machine learning algorithms such as K-Nearest Neighbors classifier help solve this problem by providing high accuracy performance. K-Nearest Neighbors is one of the machine learning algorithms used to enhance the diagnostic accuracy of the mammogram. This paper reviews some recent studies that highlight K-Nearest Neighbors accuracy, as a machine learning algorithm, in diagnosing cancer of breast.

INTRODUCTION

Machine Learning (ML), is a kind of Artificial Intelligence (AI) which allows a machine to learn by providing a set of data in order to get information via experience without excessive programming [1], [2], [3], [4], [5]. The objectives of ML are to allow machines to make predictions, do clustering, extract association rules, or make decisions from a data set. Classification strategies are most useful for the selection of approaching instances, based on certain patterns and constraints [6]. ML methods include absolute conditionality, boolean logic, and unconventional optimization approaches to make construct prediction models and patterns [7]. Depending on the data used and their availability [8]. ML consists of four stages: data collection, model selection, model training and model testing [9]. In the process of machine learning, a machine is trained with data to make a decision for similar cases [19]. It is applied in many sections e: g; object recognition, network, security, and healthcare. In the field of oncology, machine learning can be used effectively to differentiate a malignant lesion from a benign one [11]. One of the most common malignancies worldwide is breast cancer (BC), which is a leading reason of death among women [12]. BC is the common disease in women in the real world. It is a cancer that develops in female breast cells by spreading into the body's surrounding tissues [13], [14], [15], [16]. However, the mortality figures for this disease have been declining, thanks to better diagnosis and management [17]. Earlier diagnosis can improve the survival rates in BC patients [18]. There is no recent association between ML and cancerous diseases, since it was used for decades to classify many types of malignancies, including breast cancer. ML is widely accepted as the system of choice for BC pattern recognition and prediction modelling [19]. After being suspected, BC can be diagnosed using the mammogram which is a very effective tool as it can detect a breast lesion (a mass or a microcalcification) even two years before being felt by the patient [20], [21], [22], [23]. The diagnosis of such lesions, however, may be missed owing to distraction or fatigue while interpreting the mammogram report. As such, automatic classification via Computer-Aided Diagnosis (CAD) is important. This can help health care professionals to properly describe a mass or a microcalcification. It has been found that CAD can improve cancer diagnosis by up to 77% [24]. This can be done with the use of advanced techniques, including AI [25]. An algorithm was developed using K-NN, which analyzes the mammographic dataset to predict breast tissue malignancy using predefined features. A set of data is reserved for the algorithm to train, allowing the remaining values to test for accuracy [26]. K-NN is one of the easiest types of classification method, it is a supervised learning method used to diagnose cancer, heart disease [27], and hepatitis [28], [29]. K-NN is relatively simple and effective classification algorithm when compared to other algorithms [30], it is non-pragmatic algorithm i: e; does not need the assumption for distributing data [31], [32], [33]. It classifies the case study directly by the samples in the data set and thus, does not require a training process [34], [35], [36]. It's an easy supervised learning algorithm for pattern recognition. K-NN algorithm stores all cases and categorizes new cases dependent on similarity measures; searches the pattern space for the k training tuples nearest to the unknown tuples. Performance relies on the optimum number of neighbors (k) chosen, which varies from one data sample to another. After the above-mentioned points, this review will focus on the usage of the K-NN algorithm as a classifier for mammogram images and how accurate is it in identifying sinister breast lesions. In this review, we tried to search for the latest studies focusing on the efficiency of K-NN as an ML algorithm in predicting BC. In addition, a detailed explanation of the approaches, data sets, and findings- achieved in this studies-are compared. The rest of the paper is structured as follows: section 2 contains ML algorithms, section 3 consists of mammogram processing stages, section 4 contains related work of the subject, in section 5 comparison and discussion of the findings are explained, and in section 6 the review is concluded.

Machine Learning Algorithms for Classification of Breast Cancer

The main types of ML are supervised learning (SL) and unsupervised learning (USL). Briefly, SL needs training with labelled data that has inputs and target output [37]. In comparison to SL, USL does not need labelled training data, and the environment only contains inputs without desirable goals [38]. In SL, this method can be described as a problem of classification and regression. The role of classification leads to a learning process which categorizes the data into a number of finite classes [39]. In the field of regression problems, the learning function maps the data to a real-value variable. Consequently, the value of the predictive variable may be calculated for each new sample based on this method [40]. The classification model is used for discrete value problems, although the regression pattern is used to make choices on continuous value problems [41], [42]. Clustering is a typical form of USL activity in which you try to identify groups or clusters to describe data objects. On the basis of this method, each new sample may be allocated to one of the established clusters with common characteristics that they shared. Suppose, for instance, that we have collected patient records relating to breast cancer and attempt to predict whether a tumor is malignant or benign, depending on its size. The ML problem will be referred to as an assessment of the likelihood that the tumor is malignant or not (If YES=1 or NO=0) [43], [44]. The analysis of classification in statistics is broad, and there are many types of classification algorithms that you can use depending on the dataset you are dealing with. The most popular algorithms for ML such as (Naïve Bayes (NB), Support Vector Machine (SVM), K-NN, Random Forest (RF), Decision Tree (DT) [45]. As shown in Figure1. ML types.

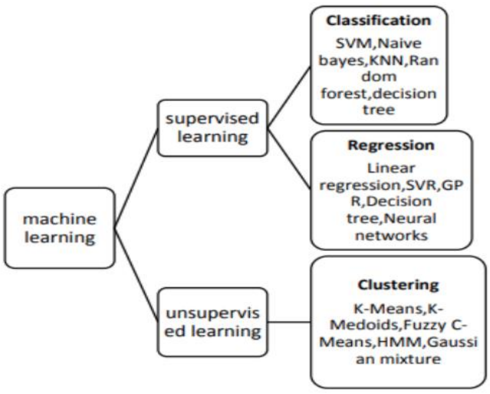

Figure 1 Types of ML [46].

The most common form of cancer among females is BC [47]. It is reported that one in eight women is at risk of developing a breast tumour in their lifetime [48]. Generally, cancer lesion is made up of a vast number of malignant cells. The rest of the breast is made up of fat, connective and lymphatic tissue [49]. Cancer cells may spread locally or to distant parts of the body, referred to as metastases [50]. Early diagnosis and prevention are essential approaches to reduce BC mortality [51]. A mammogram is a useful tool for early diagnosis of BC. Computerized processes have recently proven a fantastic tool in giving a second opinion to radiologists on the detection of BC

[52], [53]. Researchers have used CAD methods to diagnose breast tumors and to detect microcalcifications [54], [55], [56], [57]. A mammogram is a diagnostic instrument used to classify breast cancer, first developed by Bob Eagan in 1950, with a high mammogram inside an emulsion film achieving a low kVp (Peak kilovoltage). A mammogram is much more effective than a physical examination to locate abnormal masses in the breast. The categorization of image anomalies is carried out by image processing techniques to provide effective measurement of abnormalities. as shown in **Figure 2.** calcification(a), circumscribed masses (b), speculate masses(c), ill-defined masses (d), architectural distortion(e) and, asymmetry (f).

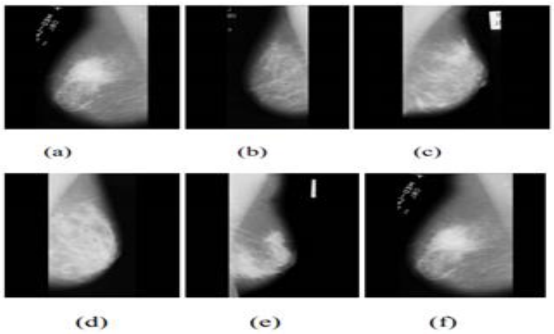

Figure 2. abnormalities commonly found on mammogram [58].

Mammogram processing stages

Masses and calcifications have been the most common anomalies for BC [59], [60]. CAD strategies by processing and analyzing mammogram images can help the mass detection and classification by the radiologist. Image processing as shown in Figure 3 includes (preprocessing, segmentation, feature extraction, feature selection and classification). First, the mammogram images are extracted from databases such as Mammographic Image Analysis Society (MIAS) database is used [90], then the mammogram images go through these stages:

1) Preprocessing: This the first step of image processing. It must be done on digitized images in order to minimize noise and improve the quality of the image. Segment the breast and improve the quality of the images.
2) Segmentation: This stage is used to classify anomalies containing suspicious regions of interest (ROIs). In general, the segmentation of images is a way of separating the image into homogeneous areas, but the unity of no two adjacent areas converts homogeneous. Segmentation strategies can be divided into two different groups (a) regional and (b) contour-based approaches [61], [62].
3) Extraction phase: Here the features are determined from the characteristics of the ROI. When choosing a feature, the best features are selected to eliminate false positives and to distinguish lesion forms.
4) Feature Selection: Is a selection of a smaller feature subset that contributes to the largest value of any classifier performance function [63].
5) Classification stage: Ultimately, the false positive reduction and the lesion classification are carried out at this stage on the basis of the characteristics selected [64]. In the application of CAD, classifiers play an important role.

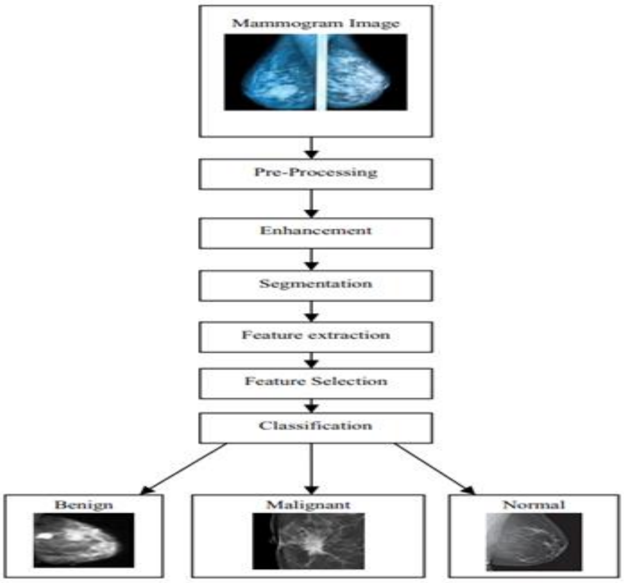

Figure 3 classification processing stages [65]

Segmentation

For mammograms, segmentation is mainly concerned with the extraction or removal of pectoral muscles, breast tumor lesions, microcalcifications, Suspected region or mass of the breast, and region of interest from the image. Clusters of microcalcification are actually tiny granular deposits of calcium that appear in a mammogram as bright spots. In dense tissues, they are difficult to notice. To increase the accuracy of classification, they are expected to be segmented from mammograms. Depending on the typical view used during mammography, the presence of pectoral muscles is also an issue in mammogram image perception. It is popular to use medieval oblique (MLO) and cardio candle (CC) views. The appearance of the pectoral muscle in the case of an MLO view is a right-angled triangle, while in the case of a CC vision, it is semi-elliptical around the breast. They are often thick with elevated contrast, and often thin with low contrast [66]. They are needed to be segmented to make a hurdle to the automated identification of lesions. Due to the complexity, shape, and homogeneity of the surrounding tissues, segmentation of the pectoral muscles is difficult. To address the limitations of conventional approaches, a large number of methods are suggested. As shown in Figure 4 visually, a shape-shifting silhouette reflects the removal of the pectoral muscles [67].

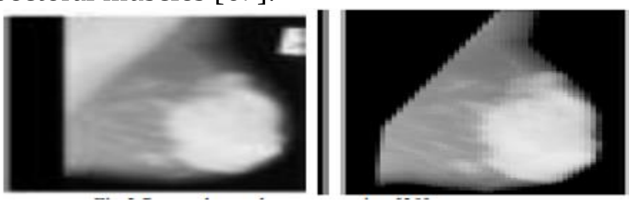

Figure 4 pectoral muscle segmentation [68]

In addition, segmentation of a region of interest and abnormal mass areas allows cancer detection and diagnosis by radiologists. Also, their isolation increases the designation's precision.

Feature extraction

We measure several features related to the geometry and texture of the boundary and its neighboring areas after segmenting the mass of the ROI. We recognize that there is a round, smooth, well-circumscribed boundary of the regular, benign mass, whereas the borderline of a malignant tumor is usually speculated, rough, and blurred [69]. Therefore, to describe the masses as benign or malignant, we may use a boundary evaluation. We investigate both geometry and texture characteristics. After segmenting the mass of the ROI, we calculate several features related to the boundary's geometry and texture and its neighboring areas. We know that the typical benign mass has a round, smooth, well-circumscribed boundary, while the boundary of the malignant tumor is typically spicules, rough, and blurred. Therefore, to describe the masses as malignant or benign, we can use a boundary examination. Consider both the properties of geometry and texture [70]. Geometry features: To explain the shape of the mass contour, the geometry features are extracted. These are calculated from the segmentation's boundary pixels. Texture features: The texture information of the region around the mass boundary also provides useful information to differentiate between malignant or benign masses, in addition to the mass contour's shape information. Hence, for the mass classification, we also use the texture features [71], [72].

K-Nearest Neighbors classifier

K-NN technique is one of the earliest and easiest ML classification algorithms [73], [74]. K-NN Classifier is a well-known way of discriminating between healthy and diseased cases after selection of features. [75], [76], [77]. However, it also generates competitive outcomes, and in some cases, when cleverly combined with prior experience, the state-of-the-art has improved considerably [78]. The K-NN rule categorizes each unknown instance in the training set by a majority marking among its nearest K-NN neighbours. Its performance is also crucially dependent on the distance metric used to define the most immediate neighbours. Most K-NN classifiers use simple Euclidean metrics to quantify the difference between examples represented as vector inputs in the absence of prior information. In addition to the conventional distance method, such as Minkowski, Chebyshev, the other suggested distance measurement formulas include Xing distance calculations [79]. The Euclidean distance is calculated as shown in the following formula.

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^{n} w_r (a_r(x_i) - a_r(x_j))^2} \qquad (1)$$

Where an example is described as a vector $x = (a1, a2, a3, \ldots, an)$, n Is the dimensionality of the input vector, that is the number of example attributes. ar Is an example rth attribute, w_r this is the weight of the rth attribute, r is from 1 to n, the smaller $d(xi, xj)$ Are the two examples more relevant. A majority vote of its nearest k neighbors must choose on the class label given to the test example.

$$y(d_i) = arg\ max \sum_{x_j \in kNN} y(x_j, c_k) \qquad (2)$$

Where d_i is an example of a test, xj is the one of the nearest neighbors to the training set, $Y(x_j, c_k)$ shows whether x_j refers to class c_k. Equation (2) indicates that a class with the majority of its members in the nearest k neighbours will be the predictor. For example, if the 5-nearest neighbor algorithm is made a classifier, a category belongs to three of the five most close neighbours of the case and the other two belongs to category Two. We can assume that the test example belongs to class one. The technique is utilized the closest neighbours algorithm if the class label of the sample is obtained only by identifying its nearest neighbors (NN) [80]. K-NN assumes that the conditional probabilities of the class are locally stable and that large dimensions gain from bias. K-NN is an incredibly versatile classification system that does not require any training data to be pre-processed. It is unwise to use the same K-NN algorithm to choose the class labels of all test examples by specifying the same number of close neighbours. The enhanced k-NN algorithm should therefore focus on finding the required k, the number of its nearest neighbours, to get its possible class label for each test example [81].

K-NN algorithm steps

1) Enter the set of data and split it into a set of training and testing.

2) From the test sets, choose an instance and measure its distance from the
training set.

3) In ascending order, list the distances.

4) The class of the instance is the most specific class of the first 3 training instances (k=3) [88].

K-NN algorithm is like the rest of the algorithms, has its benefits and drawbacks as shown in Table 1[82], [83], [84].

Table 1: Shown the benefits and drawbacks of K-NN

Benefits	Drawbacks
• The training process is speedy, and the cost is zero.	• Computationally, it's very costly.
• Simple and fast to execute.	• It's very sensitive to unrelated features
• It copes with the noise data.	• It's a lazy algorithm that needs more time to run.
• And convincing if the training data are immense.	• • It takes a considerable amount of memory that stores all the training examples.
• The algorithm is successful in computing more than one class label for an unknown instance.	• The estimated cost is high as the distance of each instance to all training tests is needed for a computer, and the value of K must be determined.

Related work

K-NN was used in BC prediction as to the classifier. In this review study, we discuss several recent works about K-NN. The kinds of related work on K-NN approaches are summarized in Table 2.

Mohamed et al. [85] used Mammograms from the Automated Mammography Screening Database (DDSM). The three kinds of extracted features are isolated from the shape of masses, texture, and wavelet features. After removing features that are previously listed, the T-test algorithm is used to select the most appropriate features. It's an easy way to find essential features. Classification is the process of defining which extra observation belongs to a set of classes, depend on a set of learning data containing observations of which the class group is identified. Finally, a distinction is made between the three classifiers are K-NN, Artificial Neural Network (ANN) and SVM. The ANN has the best performance of 98.9 %accuracy, 100 % sensitivity and 97.8 %specificity.

Diaz et al. [86] used 110 breast mammograms are taken from Mini MIAS (Mammographic Image Processing Society), a segmented image of the breast mammogram that has been through the step of pre-processing. Then, use the First-Order extraction feature (FO) to extract the feature. Since the extraction feature's value has a diversity of values, the normalization process is performed to set the resulting value to the required range, which will become the basis for the K-NN classification method. The results achieved where the Highest accuracy values are found when the value of K = 5 has an accuracy of 91.8% for cross-validation classification and K= 15 with accuracy of 91.8 % without cross-validation.

MurtiRawat et al. [87] proposed the model that trained and tested on the Wisconsin Breast Cancer Diagnosis (WDBC) data set taken from the University of California, Irvine (UCI) machine learning repository. Pre-processing of data was performed, then Principal component analysis (PCA) is use to extraction of features, and the number of components used in PCA is 17 after the analysis of the dataset. After using PCA, three models (K-NN, Logistic Regression (LR) and Ensemble Learning) are used to identify the type of tumour, i.e., benign or malignant. There are 5 ML algorithms used in Ensemble Learning: LR, K-NN, Linear Discriminant Analysis, SVM and RF Classifier. Hard voting is used in Ensemble Learning, which estimates the outcome based on the highest majority of votes. After evaluating the results using the confusion matrix, K-NN gave the accuracy of 98.60 per cent, and the Ensemble Learning Technique of 5 ML Algorithms (LR, K-NN, Linear Discriminant Analysis, SVM and RF Classifier) provided 99.30 % accuracy. From the result, the ensemble learning model achieved the best accuracy.

Amrane et al. [88] define two separate classifiers: The NB classifier and the K-NN classifier for breast cancer. Based on a study of N examples and their groups. Then, divided the data for cross-validation and testing stages. The training phase in K-NN is non-existent. Compare each new instance, every time. To classify a new model's outcome, measure the Euclidean distance between the sample and all points in the training phase. Similarly, for NB, the data collection was split into test and training sets. The training process consists first of separating the set into two separate sets: D is the presence of the tumor, and T is a set of test features and then separating the D set into two groups of malignant and benign. They considered the mean, standard deviation for every function. Set T, then set D for every class. And ended up with a description of each function and class that we're going to use for classification. A comparison between the two new implementations and an assessment by cross-validation of their accuracy. Results display that K-NN provides the highest accuracy (97.51%) with the minimum error rate compared to the NB classifier (96.19 %) on the Wisconsin Breast Cancer datasets (WBCD) datasets since their goal and challenge in the classification of BC it is to build precise and accurate classifiers.

Chaitanya et al. [89] proposed SL strategies are used to remove features that describe cancer and distinguish cancer images from regular mammogram images, a mammogram taken from DDSM. Initially, the supervised model is trained by extracting 13 characteristics from a dataset of 30 images. The extracted features of the image for test are similar to the features extracted from the training data to identify cancer tumors in the image and classify them. SVM and K-NN are used for classifying. Based on the analysis, the system is capable of producing the highest classification accuracy with K-NN. Using image processing techniques on the Matlab platform, the mechanism for detection and classification of breast cancer is suggested. The cancerous breast image and the classification results displayed in the Graphic User Interface (GUI) window can be successfully defined by the proposed mechanism. In this paper, an initial examination of the patient's symptoms could be supported by the chatbot environment advanced by using dialogue flow. The sample environment outcomes are depicted.

Htay et al. [90] used mammogram images from the MIAS database. To enhance the quality of the image's pre-processing stage, use a median filter to eliminate noise, then cropping is performed on the image. The Otsu threshold would then be used for the segmentation of the breast area on a non-uniform background. The extraction mechanism depends on the first-order statistical and Gray Level Co-occurrence Matrix (GLCM) focused textural features extraction methods, following the threshold of the images. Finally, all objects are categorized, whether they are healthy or abnormal, using the k-NN algorithm. The accuracy of the k-NN is 92%.

Laghmati et al. [91] used Wisconsin Dataset for Breast Cancer (BCWD) dataset is multivariate and contains 569 cases. The BCWD introduced CAD techniques; 32 statistical attributes between inputs and outputs are grouped together. The Neighborhood Component Analysis (NCA) mechanism is used to identify unnecessary features to eliminate as several characteristics as possible and therefore pick the most important ones. Finally, many data mining methods. k-NN, DT, Binary SVM, and AdaBoosting, are used to identify the tumour as benign or malignant. The predictive highest accuracy was for the k-NN model, the best specificity was 98.86 % for the binary SVM algorithm, and the highest predictive sensitivity was achieved for both the Adaboost models and k-NN.

Mostafa et al. [92] presented a polynomial pixel fitting technique within the Region of Interest (ROI) to obtain a original set of features that can be used to distinguish between normal and abnormal cases. These recommended features are depending on polynomial fittings for pixel strengths along the longest axes beginning from the focus point of each ROI. In this paper, as compared to many other previous works used by several feature sets, few features were used. Different methods have been used to reduce dimensionality, which could complicate the system and require a high computational cost—noting that the coefficients obtained from a polynomial with degree one are the best coefficients to classify normal from malignant ROIs using either K-NN or ANN classifiers. The highest accurate identification accuracy is 96 per cent for K-NN classifiers and 92 per cent for ANN classifiers. The suggested techniques have been extended to 50 case studies of 'BAHEYA Foundation for early detection & Treatment of BC.

Saraswathi et al. [93] used mammographic from MIAS. Then, the image is pre-processed using the median filter, and the ROI is segmented by using the Otsu threshold method. Feature extraction is achieved by using the (GLCM) ROI features that were then added to the classifier. In this CAD scheme, classifiers such as SVM and KNN were used, and performance metrics were evaluated. The accuracy of the SVM classification is 95.7%, and the sensitivity is 0.91, which is better than the KNN classifier.

Vaturi et al. [94] suggested to develop a model for detecting and correctly classifying a tumour with high precision. To do this, compare SVM with five other ML Algorithms, namely DT Classifier (CART), NB, Linear Discriminant Analysis (LDA), LR, and K-NN. ML Algorithms are known for their efficiency in the classification of data and are therefore commonly used for diagnostic purposes in the medical field. We evaluated the efficiency of SVM using precision, recall, ROC area, and accuracy estimates. The best performance was obtained by the SVM method, which resulted in the highest accuracy. Find that ML techniques provide a better understanding than other image processing techniques. ML techniques provide various probabilistic and statistical methods to identify and detect different dataset patterns. Out of these six ML algorithms, we find that SVM has the highest accuracy of 98% when applying the 3-fold cross-validation method. Then, LR and LDA followed, which gave the accuracy of 97.23 % and 95.73 %, respectively. Since SVM, LDA, and LR algorithms produce significantly more outstanding results than the others.

Bharat et al. [95] define many ML-algorithms SVM, DT CART, NB, and k-NN for the prediction of cancerous using images from WBCD. The dataset is also trained with the other algorithms: NB, K-NN, and CART, and each algorithm performs differently depending on the dataset and the parameter set. For the general strategy, the K-NN technique has achieved the best results compared with NB and CART. Additionally, SVM is an excellent predictive technique when using the Gaussian kernel is the most appropriate technique for the recurrence/non-recurrence prediction of breast cancer.

Gbenga et al. [96] evaluated the efficiency of eight (8) ML algorithms (SVM, k-NN, Radial Based Function (RBF), Simple Linear Logistic Regression Model (SLR), NB, AdaBoost, Fuzzy and DT-J48) in WBCD datasets. The comparison was made in terms of performance and efficacy of these methods using accuracy, FPR, TPR precision and F-measure to evaluate the algorithm with accuracy. SVM has the highest results with an accuracy of 97.07 %and the minimum error rate compared to RBF (96.49 %), SLR (96.78 %), NB (96.48 % &), k-NN (96.34 %), AdaBoost (96.19 %), Fuzzy to Role Induction algorithm (96.78 %) and DT-J48 (96.48 %).

Slam et al. [97] presented Compare five SL strategies called SVM, K-NN, RF, ANN, and logistic regression. Used of WBCD from UCI ML database. The performance of the research was calculated with accuracy, sensitivity, specificity, negative predictive performance, false-negative rate, false-positive rate, F1 Score, Matthews Correlation Coefficient. The findings show that the ANNs have achieved the best accuracy, Precision, and F1 scored 98.57 percent, 97.82 percent, and 0.9890 percent, respectively.

Tiwari et al. [98] demonstrate different models that are implemented as LR, SVM, K-NN, multi-layer perceptron, ANN, used the dataset from the repository of Kaggle repository. Every algorithm has been measured and compared with respect to accuracy and Precision obtained. The result showed that SVM and RF Classifier are the best predictive analyzes with an accuracy of 96.5 %. Deep learning algorithms such as CNN and ANN have been applied to improve the accuracy of predictions. The highest accuracy ANN and CNN is 99.3 % and 97.3 %, respectively.

Singh et al. [99] suggested model is intended for binary classification Images of histopathology of breast cancer. The model includes Preprocessing, extraction, training, and testing of functionality, and Classification of the publicly available BreakHis dataset. Results of the proposed model measured by Precision, Precision, remember, F-score. The accuracy is going out to be

92.3 % with an SVM cubic classifier. It is noted from the review that the magnification factor decreases the accuracy of the model. In the future, researchers proposed focusing on soft computing methods and the role of magnification in the classification model.

Mohan et al. [100] used data from the CBIS-DDSM dataset. The image is initially pre-processed, and then features for further classification are extracted from it using a 2D median filter, noise is eliminated. The features are extracted using the extraction techniques of the local binary pattern (LBP) and GLCM. After that, by using SVM and K-NN as classifiers. The result indicates the accuracy of SVM and K-NN is 96% and 100%, respectively.

Priyadharshini et al. [101] used four different classification methods such as SVM, NB, K-NN, and ANN for breast cancer detection. The dataset is imported into these algorithms, and the data is pre-processed. To train and analyze the data, the data is then divided into 8:2 ratios. Prediction is made after testing and training. The algorithm is tested in terms of accuracy using true negative, true positive, false positive and false negative values and reports on classification using the f-score, precision, and recall. The experimental findings show that the method of ANN offers greater accuracy, lesser error, and higher f-score than other algorithms.

Table 2: main characteristic of reviewed study

Reference	Year	Methods	Dataset	Feature selection	Feature extraction	Result	Accuracy
[93]	2017	• SVM • K-NN	MIAS	–	GLCM	The classification accuracy is higher for SVM classifier than the K-NN.	SVM: 95.7%
[96]	2017	• RBF • SVM • SLR • NB • KNN, AdaBo	699 instances-UCI	Probability technique	–	SVM has the greatest result with the highest accuracy.	SVM: 97.07 %

		ost, Fuzzy and DT-J48					
[88]	2018	• K-NN • NB	683 samples - UCI	Probability technique	–	K-NN achieved higher efficiency. However, even NB has good accuracy, if the dataset is larger, the KNN's for running time will increase.	K-NN: 97.51 %
[85]	2018	• ANN • SVM • K-NN	250 mammogram- DDSM	T-test algorithm	Wavelet transform	ANN classifier proved to be an efficient technique when compared to other alternatives in the literature.	ANN: 98.9%
[90]	2018	• K-NN	120 images- Mini-MIAS	–	GLCM	K-NN classifier made a high accurac	K-NN: 92%

						y prediction.	
[95]	2018	• SVM • DT (CART), NB, k-NN	357 data set-WBCD	features that are calculated from a digitized image of a fine needle aspiration (FNA) biopsy		SVM has the best result than other algorithms.	SVM: 92%
[86]	2019	• KNN	110 mammogram-Mini-MIAS	–	Texture analysis	The result shows that the classification without cross-validation achieved the highest accuracy when k is fifteen but from the efficiency side, the highest accuracy when using cross-validation when K is five.	K-NN: 91.8% (cross-validation)
[89]	2019	• SVM • K-NN	30 images-DDSM	–	Contrast Correlation	The K-NN algorithm	K-NN: 97%

						achieved a higher accuracy result in differentiating normal and cancerous breast image	
[92]	2019	• K-NN • ANN	50 CESM images BAHAYA foundation	–	Curve fitting	The accuracy results for KNN and ANN were 96% and 92% respectively.	K-NN :96% ANN: 92%
[87]	2020	• K - NN • LR • Ensemble Learning	569 instances - WDBC	Statical technique Correlation coefficient	PCA	Ensemble Learning technique of 5 machine learning algorithms gave the highest performance.	Ensemble learning: 99.30 %
[91]	2020	• K-NN • DT • B	569 instances-BCWD	NCA	–	k-NN is the best model in term	K-NN: 99.12 %

		inary SVM • Adaboost				of both accuracy and sensitivity, BSVM is the best model with the highest specificity	
[94]	2020	• SVM • DT • NB • LR • LDA • K-NN	569 instances-UCI	–	PCA	The best performance was achieved by the SVM method resulting in the highest accuracy.	SVM: 98% LR: 97.23 % LDA: 95.73 %

COMPARISON AND DISCUSSION

Table 2 introduced different authors using the different dataset, K-NN with other algorithms, result and accuracy of the technique. Using of K-NN with other techniques in previous studies will be discussed in more details in this section. Mohamed et al. demonstrated that the performance of the ANN Classifier is improved with the increase in the number of features that reach the highest precision than K-NN accuracy. The approach discussed in [87] obtained an accuracy of 98.60 % using K-NN, 97.90% using LR and 99.30% using Ensemble Learning. Data which have applied to the database helped to improve the training of ML models and to function more specifically and also notified the interactions among the attributes. Researchers in [95] noted that the performance of the SVM is greatly enhanced by using standardized data as input to the SVM classifier. Since SVM, LR and LDA algorithm generate a massively higher result than the K-NN, it is important to examine these algorithms and try to achieve better results. Furthermore, ML methods have been used in [96] to identify and test the dataset and used the 10-fold cross-validation test, which is the method used to analyze the predictive models that split the original set into a testing sample to train the algorithm, and the test set for its evaluation. It is because k-NN is a lazy learner, and not much was learned during the training process, compared to what can be learned in other algorithms that generate their models. The k-NN is referred to as a lazy algorithm since the training data phase is not used for generalization. Simply stated, there is no obvious stage of training, and, if it occurs, it is extremely small. The efficiency of K-NN is, therefore, less. Similarly, in [86] the K-NN accuracy is 91.8% after character extraction using six statistical measurements of the first-order method, namely mean, smoothness, standard deviation, third moment, uniformity and entropy, the K-NN image classification procedure will operate well. In order to achieve high precision values for the training and classification process, the cross-validation strategy used in k-fold cross-validation with a fold or cv is five. Cross-validation is a mathematical technique; it is typically used to verify and validate learning algorithms or models by partitioning data into a learning set to train the model and test set to evaluate it. K-fold cross-validation is the basic method, one of the K partitions used as a validation package. There are more complex ways of cross-validation. Even after the extraction value of the function had been normalized, and the following is the distribution of the values of each feature shown on the basis of the mean, minimum and maximum values. In order to improve K-NN performance, checking with various extraction features may also be considered.

On the other hand, some researchers have suggested ways to resolve the limitations of K-NN problems, K-NN achieved the best accuracy results when compared to other algorithms. It is observed in [89] that SVM and K-NN have been used to separate cancer images from normal images. In order to improve K-NN performance Checking with various extraction features may also be considered seeing the ability of the classification results with the K-NN classification processor with other techniques. In addition, [90] compared first-order statistical features and (GLCM) feature extraction and k-NN classification for an early-stage cancer detection system, a 92% accuracy rate was achieved with K-NN. Even more, it is better to add more extracted features and to apply the selection function in order to get a more precise BC classification. In [91] k-NN has proved its strength in the classification of the severity of BC thanks to the use of the NCA technique for the selection of features. In [92] also, the used features based on polynomial fitting for values of pixel intensities along the long axes starting from the central focus of each ROI, there was no need for features of reduction strategies. And because of this, the cost of computing is very low although the K-NN has high accuracy values. In [88], it can be clearly observed that K-NN achieved greater performance than NB. While NB has great precision if the dataset is wider, K-NN execution time would increase. However, there are limitations in the K-NN algorithm itself. In recent years, it has been developed primarily in two respects first, the combination of K-NN and other technologies, and the features of the selection technique may provide a tangible effect on the precision of the K-NN for breast cancer prediction in general.

CONCLUSION

Cancer was known to be the leading cause of death in people, especially female, which is essentially caused by the abnormal development of cancer cells. This paper presented a review of the classification algorithms used to breast cancer by implementing the K-NN algorithm. Each algorithm used a different classification technique, and the dataset mostly used differed among various studies. To sum it up, the implementation of the using K-NN is relatively easy. In terms of accuracy, K-NN gave the most accurate prediction (99.12%). For breast cancer diagnostic by using various algorithms in the future offers a much more effective and reliable method for identifying the illness.

ACKNOWLEDGEMENT

The authors would like to note that, with all its financial resources and encouragement, Duhok Polytechnic University supported this report.

REFERENCES

E. D. Michie, D. J. Spiegelhalter, And C. C. Taylor, “Machine Learning, Neural And Statistical Classification,” Proceeding, 1994.

Bargarai, F., Abdulazeez, A., Tiryaki, V., & Zeebaree, D. (2020). Management Of Wireless Communication Systems Using Artificial Intelligence-Based Software Defined Radio.

D. Abdulqader, A. Mohsin Abdulazeez, And D. Zeebaree, “Machine Learning Supervised Algorithms Of Gene Selection: A Review,” Apr. 2020.

D. Maulud And A. M. Abdulazeez, “A Review On Linear Regression Comprehensive In Machine Learning,” J. Appl. Sci. Technol. Trends, Vol. 1, No. 4, Pp. 140–147, 2020.

Jahwar, A. F., & Abdulazeez, A. M. (2020). Meta-Heuristic Algorithms For K-Means Clustering: A Review. Palarch's Journal Of Archaeology Of Egypt/Egyptology, 17(7), 12002-12020.

Al-Shargabi, B., & Al-Shami, F. A. (2019, December). An Experimental Study For Breast Cancer Prediction Algorithms. In Proceedings Of The Second International Conference On Data Science, E-Learning And Information Systems (Pp. 1-6).

Jf Mccarthy, M.K., Pe Hoffman, "Applications Of Machine Learning And High- Dimensional Visualization In Cancer Detection, Diagnosis, And Management", Ann N Y Acad Sci, Vol.62, Pp. 10201259, 2004.

Ac. Tan, D. Gilbert, "Ensemble Machine Learning On Gene Expression Data For Cancer Classification.

S. Gokhale., "Ultrasound Characterization Of Breast Masses", The Indian Journal Of Radiology &Imaging, Vol. 19, Pp. 242-249, 2009.

P. Louridas, And C. Ebert, “Machine Learning,” Ieee Softw., Vol. 33, No. 5, Pp. 110–115, 2016.

T. Jinshan, R.R., X. Jun, I. El Naqa, Y. Yongyi, "Computer-Aided Detection And Diagnosis Of Breast Cancer With Mammography: Recent Advances", Information Technology In Biomedicine. Ieee, Vol. 13, Pp. 236-251, 2009.

U.S. Cancer Statistics Working Group. United States Cancer Statistics: 1999–2008 Incidence And Mortality Web-Based Report. Atlanta (Ga): Department Of Health And Human Services, Centers For Disease Control.

Wan, K., & Zunaidi, I. (2017). An Efficient Data Mining Approaches For Breast Cancer Detection And Segmentation In Mammogram. Jour Of Adv Research In Dynamical & Control Systems.

Shofwatul'Uyun, L. C. (2018). Feature Selection Mammogram Based On Breast Cancer Mining. International Journal Of Electrical And Computer Engineering (Ijece), 8(1), 60-69.

Breast Cancer Q&A/Facts And Statistics (Http://Www.Komen.Org/Bci/Bhealth/Qa/Q_And_A.Asp)

Jerez-Aragone ´S Jm, Gomez-Ruiz Ja, Ramos-Jimenez G, Munozperez J, Alba-Conejo E. A Combined Neural Network And Decision Trees Model For Prognosis Of Breast Cancer Relapse. Artif Intell Med.

B. E. Sirovich And H. C. Sox, Jr., "Breast Cancer Screening," Surgical Clinics N. Amer., Vol. 79, No. 5, Pp. 961–990, Oct. 1, 1999, 10.1016/ S0039-6109(05)70056-6, 0039-6109.

De Oliveira, J.P.S., Conci, A., Prez, M.G., Andaluz, V.H.: Segmentation Of Infrared Images: A New Technology For Early Detection Of Breast Diseases. In: 2015 Ieee International Conference On Industrial Technology (Icit), Pp. 1765–1771 (2015).

Yue, W., Wang, Z., Chen, H., Payne, A., & Liu, X. (2018). Machine Learning With Applications In Breast Cancer Diagnosis And Prognosis. Designs, 2(2), 13.

Isource National Breast Cancer Centre, 2011. Clinical Practice Guidelines For The Management Of Early Breast Cancer, National Health & Medical Research Council, Australia.

Radiologyinfo, 2014. Mammography, Available Online At: Www.Radiologyinfo.Org/En/Info.Cfm?Pg=Mammo, Accessed On 12 November 2014

L. Tabár And P. Dean, "A New Era In The Diagnosis And Treatment Of Breast Cancer," J. Breast, Vol. 16, Pp. S2–S4, 2010.

Arbach, L., Reinhardt, J. M., Bennett, D. L., & Fallouh, G. (2003, May). Mammographic Masses Classification: Comparison Between Backpropagation Neural Network (Bnn), K Nearest Neighbors (Knn), And Human Readers. In Ccece 2003-Canadian Conference On Electrical And Computer Engineering. Toward A Caring And Humane Technology (Cat. No. 03ch37436) (Vol. 3, Pp. 1441-1444). Ieee.

Birdwell, R.L., Ikeda, D.M., O'shaughnessy, K.F., Sickles, E.A., 2001. Mammographic Characteristics Of 115 Missed Cancers Later Detected With Screening Mammography And The Potential Utility Of Computer-Aided Detection. Radiology, Volume 219(1), Pp. 192□202.

S. A. Korkmaz, And M. Poyraz, "A New Method Based For Diagnosis Of Breast Cancer Cells From Microscopic Images: Dwee—Jht," J. Med. Syst., Vol. 38, No. 9, P. 92, 2014.

Rodriguez, V., Sharma, K., & Walker, D. (2018). Breast Cancer Prediction With K-Nearest Neighbor Algorithm Using Different Distance Measurements. Software Engineering Project (Swen 670), University Of Maryland, University College, Usa.

O. Terrada, B. Cherradi, A. Raihani, And O. Bouattane, "Atherosclerosis Disease Prediction Using Supervised Machine Learning Techniques," In 2020 1st International Conference On Innovative Research In Applied Science, Engineering And Technology (Iraset), Meknes, Morocco, Apr. 2020, Pp. 1–5, Doi: 10.1109/Iraset48871.2020.9092082.

E. M. F. El Houby, "A Survey On Applying Machine Learning Techniques For Management Of Diseases," J Appl Biomed, Vol. 16, No. 3, Pp. 165– 174, Aug. 2018, Doi: 10.1016/J.Jab.2018.01.002.

S. Medjahed, T. Saadi, And A. Benyettou, "Breast Cancer Diagnosis By Using K-Nearest Neighbor With Different Distances And Classification Rules," International Journal Of Computer Applications, 2013, Vol. 62, No. 1, Pp. 0975 – 8887.

Jabbar Ma, Deekshatulu Bl, Chandra P. "Classification Of Heart Disease Using K-Nearest Neighbor And Genetic Algorithm". International Conference On Computational Intelligence: Modeling Techniques And Applications (Cimta).2013; 10:85-94.

Sarkar M, Leong T-Y. Application Of K-Nearest Neighbors Algorithm On Breast Cancer Diagnosis Problem. In: Proceedings Of The Amia Symposium. American Medical Informatics Association; 2000. P. 759.

Gou J, Du L, Zhang Y, Xiong T. A New Distanceweighted K-Nearest Neighbor Classifier. J Inf Comput Sci. 9(6):1429–1436.

Jain R, Mazumdar J. A Genetic Algorithm Based Nearest Neighbor Classification To Breast Cancer Diagnosis. Australas Phys Eng Sci Med. 2003;26(1):6–11.

Senturk Zk, Kara R. Breast Cancer Diagnosis Via Data Mining: Performance Analysis Of Seven Different Algorithms. Comput Sci Eng. 2014;4(1):35.

Odajima K, Pawlovsky Ap. A Detailed Description Of The Use Of The Knn Method For Breast Cancer Diagnosis. In: Biomedical Engineering And Informatics (Bmei), 2014 7th International Conference. Ieee; 2014 May 27. P. 688–692.

Medjahed Sa, Saadi Ta, Benyettou A. Breast Cancer Diagnosis By Using K-Nearest Neighbor With Different Distances And Classification Rules. Int J Comput Appl. 2013;62(1).

Zebari, D. A., Zeebaree, D. Q., Abdulazeez, A. M., Haron, H., & Hamed, H. N. A. (2020). Improved Threshold Based And Trainable Fully Automated Segmentation For Breast Cancer Boundary And Pectoral Muscle In Mammogram Images. Ieee Access, 8, 203097-203116.

Qiu, J., Wu, Q., Ding, G., Xu, Y., & Feng, S. (2016). A Survey Of Machine Learning For Big Data Processing. Eurasip Journal On Advances In Signal Processing, 2016(1), 67.

Zeebaree, D. Q., Abdulazeez, A. M., Zebari, D. A., Haron, H., & Hamed, H. N. A. Multi-Level Fusion In Ultrasound For Cancer Detection Based On Uniform Lbp Features.

Fatima, M., & Pasha, M. (2017). Survey Of Machine Learning Algorithms For Disease Diagnostic. Journal Of Intelligent Learning Systems And Applications, 9(01), 1.

N. K. Chauhan And K. Singh, "A Review On Conventional Machine Learning Vs Deep Learning," In 2018 International Conference On Computing, Power And Communication Technologies (Gucon), 2018, Pp. 347–352.

S. B. Kotsiantis, I. Zaharakis, And P. Pintelas, “Supervised Machine Learning: A Review Of Classification Techniques,” Emerg. Artif. Intell. Appl. Comput. Eng., Vol. 160, Pp. 3–24, 2007.

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.

Marshland, S. (2009) Machine Learning An Algorithmic Perspective. Crc Press, New Zealand, 6-7.

Issa, Adel Sabry, And Adnan Mohsin Abdulazeez Brifcani. "Intrusion Detection And Attack Classifier Based On Three Techniques: A Comparative Study." Engineering And Technology Journal 29, No. 2 (2011): 386-412.

Abdulqader, D. M., Abdulazeez, A. M., & Zeebaree, D. Q. (2020). Machine Learning Supervised Algorithms Of Gene Selection: A Review. Machine Learning, 62(03).

G. Lalli, N. Manikandaprabu, D. Kalamani And C. N. Marimuthu, "A Development Of Knowledge-Based Inferences System For Detection Of Breast Cancer On Thermogram Images," 2014 International Conference On Computer Communication And Informatics, Coimbatore, 2014, Pp. 1-6, Doi: 10.1109/Iccci.2014.6921743.

Zeebaree, D. Q., Haron, H., Abdulazeez, A. M., & Zebari, D. A. (2019, April). Machine Learning And Region Growing For Breast Cancer Segmentation. In 2019 International Conference On Advanced Science And Engineering (Icoase) (Pp. 88-93). Ieee.

[49] Breast Cancer Facts & Figures, 2009-2010, American Cancer Society, Inc.

B. Stewart And C.P. Wild, World Cancer Report 2014, International Agency For Research On Cancer, Who, 2014.

Nicolò, C., Périer, C., Prague, M., Bellera, C., Macgrogan, G., Saut, O., & Benzekry, S. (2020). Machine Learning And Mechanistic Modeling For Prediction Of Metastatic Relapse In Early-Stage Breast Cancer. Jco Clinical Cancer Informatics, 4, 259-274.

Nithya, R., & Santhi, B. (2011). Classification Of Normal And Abnormal Patterns In Digital Mammograms For Diagnosis Of Breast Cancer. International Journal Of Computer Applications, 28(6), 21-25.

Kelly, K. M., Dean, J., Comulada, W. S., & Lee, S. J. (2010). Breast Cancer Detection Using Automated Whole Breast Ultrasound And Mammography In Radiographically Dense Breasts. European Radiology, 20(3), 734-742.

E. Claridge And J. Richer, "Characterization Of "Nographic Lesions," In 2nd Lnremolionul Workshop On Digital Mammography, York, England, July 1994, Pp. 241-250.

L. Arbach. N. Ai-Mudarres, And M. Ammar, "Computerized Detection Of Breast -Se$ From Mammograms:' The Thirry Sevenrh Week Of Science, Damascus. Syria. Junc 1997.

Z. R. Yang And R. G. Harrison, "Detecting False Benign In Breast Cancer Diagnosis," Neural Neworks, Ieee, Vol. 3, Pp. 655458. July 2000.

N. Pandey And J. Sivaswamy, "Fuzzy Logic Bascd Microcalcification Detection," In Ieee Conference On Neural Nework, Vol. 2, Dec 2000, Pp. 662471.

Kamalakannan, J., Thirumal, T., Vaidhyanathan, A., & Mukeshbhai, K. D. (2015, March). Study On Different Classification Technique For Mammogram Image. In 2015 International Conference On Circuits, Power And Computing Technologies [Iccpct-2015] (Pp. 1-5). Ieee.

D.W.Kufe,R.E. Pollock,R.R.Weichselbaum, Et Al.,Holland-Freicancermedicine, 6th Edition, Bc Decker Inc., Hamilton, 2003.

Maitra, I. K., Nag, S., & Bandyopadhyay, S. K. (2011). Identification Of Abnormal Masses In Digital Mammography Images. International Journal Of Computer Graphics, 2(1), 17-30.

Elmoufidi, A., El Fahssi, K., Jai-Andaloussi, S., Sekkaki, A., Gwenole, Q., & Lamard, M. (2017). Anomaly Classification In Digital Mammography Based On Multiple-Instance Learning. Iet Image Processing, 12(3), 320-328.

R. Garg, ‘7 Types Of Classification Algorithms’, 2018.

Adel Sabry Eesa, Zeynep Orman, Adnan Mohsin Abdulazeez Brifcani, “A Novel Feature-Selection Approach Based On The Cuttlefish Optimization Algorithm For Intrusion Detection Systems”, Expert Systems With Applications, Pp. 2670– 2679, 2015.

Ali, Y., & Hamed, S. (2015). Early Breast Cancer Detection Using Mammogram Images: A Review Of Image Processing Techniques. Biosciences Biotechnology Research Asia, 12(1), 225-234.

Meenalochini, G., & Ramkumar, S. (2020). Survey Of Machine Learning Algorithms For Breast Cancer Detection Using Mammogram Images. Materials Today: Proceedings.

[66] Abdulazeez, A. M., Zeebaree, D. Q., & Abdulqader, D. M. (2020). Wavelet Applications In Medical Images: A Review. Transformation (Dwt), 21, 22.

Dabass, J., Arora, S., Vig, R., & Hanmandlu, M. (2019, January). Segmentation Techniques For Breast Cancer Imaging Modalities-A Review. In 2019 9th International Conference On Cloud Computing, Data Science & Engineering (Confluence) (Pp. 658-663). Ieee.

Mughal, B., Muhammad, N., Sharif, M., Rehman, A., & Saba, T. "Removal Of Pectoral Muscle Based On The Topographic Map And Shape-Shifting Silhouette." Bmc Cancer18.1 (2018): 778.

N. Mudigonda, R. Rangayyan, And J. Desautels, “Gradient And Texture Analysis For The Classification Of Mammographic Masses,” Ieee Trans. Med. Imag., Vol. 19, No. 10, Pp. 1032–1043, Oct. 2000.

X. Liu, J. Liu, D. Zhou, And J. Tang, “A Benign And Malignant Mass Classification Algorithm Based On An Improved Level Set Segmentation And Texture Feature Analysis,” In Proc. 4th Icbbe, 2010, Pp. 1–4.

Zebari, R., Abdulazeez, A., Zeebaree, D., Zebari, D., & Saeed, J. (2020). A Comprehensive Review Of Dimensionality Reduction Techniques For Feature Selection And Feature Extraction. Journal Of Applied Science And Technology Trends, 1(2), 56-70.

Eesa, A. S., Brifcani, A. M. A., & Orman, Z. (2015). A New Dids Design Based On A Combination Feature Selection Approach. International Journal Of Computer And Information Engineering, 9(8), 1921-1925.

R.O. Duda And P.E. Hart, “Pattern Classification And Scene Analysis,” John Wiley & Sons, New York, 1973.

S.L. Sun, “Ensembles Of Feature Subspaces For Object Detection,” Lecture Notes In Computer Science, Vol. 5552, 2009, Pp. 996-1004.

Shahari, S., Wakankar, A.: Color Analysis Of Thermograms For Breast Cancer Detection. In: 2015 International Conference On Industrial Instrumentation And Control (Icic), Pp. 1577–1581 (2015).

Mejia, T.M., Prez, M.G., Andaluz, V.H., Conci, A.: Automatic Segmentation And Analysis Of Thermograms Using Texture Descriptors For Breast Cancer Detection. In: 2015 Asia-Pacific Conference On Computer Aided System Engineering, Pp. 24–29 (2015).

Milosevic, M., Jankovic, D., Peulic, A.: Thermography Based Breast Cancer Detection Using Texture Features And Minimum Variance Quantization. Excli J. 13, 1204–1215 (2014).

S. Belongie, J. Malik And J. Puzicha, “Shape Matching And Object Recognition Using Shape Contexts,” Ieee Transactions On Pattern Analysis And Machine Intelligence (Pami), Vol. 24(4), 2002, Pp. 509-522. [111] P.Y. Simard, Y. Lecun And J. Decker, “Efficient Pattern Recognition Using A New Transformation Distance,” In Advances In Neural Information Processing Systems, Vol. 6, 1993, Pp. 50-58.

Xing Ep, Jordan Mi, Russell S, & Ng Ay (2002) Distance Metric Learning With Application To Clustering With Side-Information. In Advances In Neural Information Processing Systems (Pp. 505–512).

B. Li, S. Yu And Q. Lu, “An Improved K-Nearest Neighbor Algorithm For Text Categorization,” Proceedings Of The 20th International Conference On Computer Processing Of Oriental Languages, China: Shenyang, 2003.

Sun, S., & Huang, R. (2010, August). An Adaptive K-Nearest Neighbor Algorithm. In 2010 Seventh International Conference On Fuzzy Systems And Knowledge Discovery (Vol. 1, Pp. 91-94). Ieee.

Bhatia, N., Et Al.(2010). Survey Of Nearest Neighbor Techniques. Arxiv Preprint Arxiv:1007.0085.

K. P. Soman, S. Diwakar, And V. Ajay, Data Mining: Theory And Practice [With Cd]. Phi Learning Pvt. Ltd., 2006.

Mynavathi, R., Bhuvaneswari, V., Karthikeyan, T., & Kavina, C. (2016). K Nearest Neighbor Classifier Over Secured Perturbed Data. In 2016 World Conference On Futuristic Trends In Research And Innovation For Social Welfare (Startup Conclave) (Pp. 1-4). Ieee.

Mohamed, B. A., & Salem, N. M. (2018, March). Automatic Classification Of Masses From Digital Mammograms. In 2018 35th National Radio Science Conference (Nrsc) (Pp. 495-502). Ieee.

Diaz, R. A. N., Swandewi, N. N. T., & Novianti, K. D. P. (2019, August). Malignancy Determination Breast Cancer Based On Mammogram Image With K-Nearest Neighbor. In 2019 1st International Conference On Cybernetics And Intelligent System (Icoris) (Vol. 1, Pp. 233-237). Ieee.

Murtirawat, R., Panchal, S., Singh, V. K., & Panchal, Y. (2020, July). Breast Cancer Detection Using K-Nearest Neighbors, Logistic Regression And Ensemble Learning. In 2020 International Conference On Electronics And Sustainable Communication Systems (Icesc) (Pp. 534-540). Ieee.

Amrane, M., Oukid, S., Gagaoua, I., & Ensari, T. (2018, April). Breast Cancer Classification Using Machine Learning. In 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (Ebbt) (Pp. 1-4). Ieee.

Chaitanya, V., Killedar, S. M., Revankar, D., & Pushpa, M. S. (2019, May). Recognition And Prediction Of Breast Cancer Using Supervised Diagnosis. In 2019 4th International Conference On Recent Trends On Electronics, Information, Communication & Technology (Rteict) (Pp. 1436-1441). Ieee.

Htay, T. T., & Maung, S. S. (2018, September). Early Stage Breast Cancer Detection System Using Glcm Feature Extraction And K-Nearest Neighbor (K-Nn) On Mammography Image. In 2018 18th International Symposium On Communications And Information Technologies (Iscit) (Pp. 171-175). Ieee.

Laghmati, S., Cherradi, B., Tmiri, A., Daanouni, O., & Hamida, S. (2020, September). Classification Of Patients With Breast Cancer Using Neighbourhood Component Analysis And Supervised Machine Learning Techniques. In 2020 3rd International Conference On Advanced Communication Technologies And Networking (Commnet) (Pp. 1-6). Ieee.

Mostafa, S., Mubarak, R., El-Adawy, M., Ibrahim, A. F., Gomaa, M. M., & Kamal, R. M. (2019, February). Breast Cancer Detection Using Polynomial Fitting Applied On Contrast Enhanced Spectral Mammography. In 2019 International Conference On Innovative Trends In Computer Engineering (Itce) (Pp. 11-16). Ieee.

Saraswathi, D., & Srinivasan, E. (2017, January). Performance Analysis Of Mammogram Cad System Using Svm And Knn Classifier. In 2017 International Conference On Inventive Systems And Control (Icisc) (Pp. 1-5). Ieee.

Ivaturi, A., Singh, A., Gunanvitha, B., & Chethan, K. S. (2020, June). Soft Classification Techniques For Breast Cancer Detection And Classification. In 2020 International Conference On Intelligent Engineering And Management (Iciem) (Pp. 437-442). Ieee.

Bharat, A., Pooja, N., & Reddy, R. A. (2018, October). Using Machine Learning Algorithms For Breast Cancer Risk Prediction And Diagnosis. In 2018 3rd International Conference On Circuits, Control, Communication And Computing (I4c) (Pp. 1-4). Ieee.

Gbenga, D. E., Christopher, N., & Yetunde, D. C. (2017). Performance Comparison Of Machine Learning Techniques For Breast Cancer Detection. Nova, 6(1), 1-8.

Islam, M. M., Haque, M. R., Iqbal, H., Hasan, M. M., Hasan, M., & Kabir, M. N. (2020). Breast Cancer Prediction: A Comparative Study Using Machine Learning Techniques. Sn Computer Science, 1(5), 1-14.

Tiwari, M., Bharuka, R., Shah, P., & Lokare, R. (2020). Breast Cancer Prediction Using Deep Learning And Machine Learning Techniques. Available At Ssrn 3558786.

Singh, S., & Kumar, R. (2020, February). Histopathological Image Analysis For Breast Cancer Detection Using Cubic Svm. In 2020 7th International Conference On Signal Processing And Integrated Networks (Spin) (Pp. 498-503). Ieee.

Mohan, P. (2020, August). A Comparison Between Knn And Svm For Breast Cancer Diagnosis Using Glcm Shape And Lbp Features. In 2020 Third International Conference On Smart Systems And Inventive Technology (Icssit) (Pp. 1058-1062). Ieee.

Priyadharshini, R., & Kavithamani, A. Comparion Of Machine Learning Algorithms For Breast Cancer Detection.