BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

Journal of Oncology Research
https://ojs.bilpublishing.com/index.php/jor

REVIEW

# Machine Learning Algorithms for Breast Cancer Diagnosis: Challenges, Prospects and Future Research Directions

**Rebecca Nyasuguta Arika*** **Agnes Mindila** **W. Cheruiyot**

School of Computing and Information Technology, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

ARTICLE INFO

*Article history*
Received: 19 August 2022
Revised: 11 October 2022
Accepted: 21 October 2022
Published Online: 2 November 2022

*Keywords*:
Algorithm
Quantum computing
Machine learning
Breast cancer
Prediction

ABSTRACT

Early diagnosis of breast cancer does not only increase the chances of survival but also control the diffusion of cancerous cells in the body. Previously, researchers have developed machine learning algorithms in breast cancer diagnosis such as Support Vector Machine, K-Nearest Neighbor, Convolutional Neural Network, K-means, Fuzzy C-means, Neural Network, Principle Component Analysis (PCA) and Naive Bayes. Unfortunately these algorithms fall short in one way or another due to high levels of computational complexities. For instance, support vector machine employs feature elimination scheme for eradicating data ambiguity and detecting tumors at initial stage. However this scheme is expensive in terms of execution time. On its part, k-means algorithm employs Euclidean distance to determine the distance between cluster centers and data points. However this scheme does not guarantee high accuracy when executed in different iterations. Although the K-nearest Neighbor algorithm employs feature reduction, principle component analysis and 10 fold cross validation methods for enhancing classification accuracy, it is not efficient in terms of processing time. On the other hand, fuzzy c-means algorithm employs fuzziness value and termination criteria to determine the execution time on datasets. However, it proves to be extensive in terms of computational time due to several iterations and fuzzy measure calculations involved. Similarly, convolutional neural network employed back propagation and classification method but the scheme proves to be slow due to frequent retraining. In addition, the neural network achieves low accuracy in its predictions. Since all these algorithms seem to be expensive and time consuming, it necessary to integrate quantum computing principles with conventional machine learning algorithms. This is because quantum computing has the potential to accelerate computations by simultaneously carrying out calculation on many inputs. In this paper, a review of the current machine learning algorithms for breast cancer prediction is provided. Based on the observed shortcomings, a quantum machine learning based classifier is recommended. The proposed working mechanisms of this classifier are elaborated towards the end of this paper.

*Corresponding Author:
Rebecca Nyasuguta Arika,
School of Computing and Information Technology, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya;
*Email: arikarebe@gmail.com*

DOI: https://doi.org/10.30564/jor.v5i1.4977

Copyright © 2023 by the author(s). Published by Bilingual Publishing Co. This is an open access article under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License. (https://creativecommons.org/licenses/by-nc/4.0/).

## 1. Introduction

Cancer is regarded as a collection of diseases and as such, any region of the body might be affected by cancerous cell development. For instance normal cells may become crowded as the cancerous growth spreads throughout the body making it difficult for the body to operate normally. According to Al-Azzam and Shatnawi [1], cancer is a life threatening disease that has claimed many lives. As explained by Rahman et al. [2], breast cancer is malignant tumor that begins in the cells of the breast which afterwards spreads to the surrounding tissues. As Ara et al. [3] discuss, tumors and lumps are as a result of cancer growth. However, some anomalies may not be harmful. For one to determine whether tumor or lump is cancerous, some samples have to be taken by pathologist for examination under the microscope. If the result proves to be cancerous, then it is regarded as malignant tumor and if the result proves to be non-cancerous, then it is regarded as benign tumor. As explained by Thalor et al. [4], cancer can be of different types, such as leukemia (blood related cancer), prostate cancer, lung cancer and breast cancer.

Breast cancer is a life threatening disease which can be cured if detected early. It is a complex disease involving multiple data types which may represent obstacle for clinical and radiological diagnosis more specifically at early stages [5,6]. Breast cancer is regarded as one of the primary diseases behind the loss of many lives, more specifically women globally [7]. It is actually the second major cause of death after lung cancer [8]. It normally occurs in both men and women, although it is more prevalent in women. Breast cancer begins in the milk duct before spreading to the areola where some have the origins in the organs responsible for producing milk [9]. The causative agent of breast cancer is still under research, although some risk factors associated with this disease are well known and include age, gene, obesity, taking birth control pills and smoking [10]. The malignant tumor may begin in the cells of the breast, which then spreads to the surrounding tissues [11-13]. As explained by Rasool et al. [14], normal cells in the breast and other parts of the body grow and divide to form new cells as they are needed. When these normal cells grow old and get damaged, they die and new cells take their place [15]. However sometimes this process goes wrong. For instance, if new cells form when the body does not need them and old or damaged cells do not die as they should, this builds up extra cells often forming mass of tissues called a lump, growth or tumor [5]. When there is an abnormal growth of cells in the grand that produce milk (lobules), then this is regarded as breast cancer [16]. Therefore tumor in the breast can be benign (not cancer) or malignant (cancer).

As discussed by Srinivas et al. [13], cancer can be classified into normal, benign and malignant tumors. It is also grouped into five stages (0-IV) depending on the size of the tumor and these stages can be identified as invasive and non-invasive. According to GLOBOCAN 2020 estimates of cancer incidence and mortality produced by international agency for research on cancer worldwide, an estimated 19.3 million are new cancer cases [17]. Among these cases, female breast cancer has surpassed lung cancer as the most commonly diagnosed cancer with an estimation of 2.3 million new cases (11.7%) followed by lung cancer (11.4%), colorectal (10.0%), prostate (7.3%) and stomach (5.6%) cancers. However GLOBOCAN 2020 estimates breast cancer mortality rate to be 6.9%. The major reason for the rise of mortality rates in breast cancer and related issues is misinterpretations of radiologists in recognizing suspicious lesions [18]. This is due to technical issues in imaging qualities which increases the false positive and negative ratio.

As explained by Sung et al. [17], the detection of lymph node metastasis in pre-operative stages can sensibly optimize care quality and efficiency as well as patient safety. However, those not affected by clinical or radiological examinations are said to be clinically negative. Moreover, among those cases, false negative patients are included especially when characterized as early stage tumors [19]. This presents some difficulties for pathologists in decision making. This is specifically the case in the choice of features, which have a significant role in the prediction process [20].

Since the root cause of the breast cancer remains unclear, early detection and diagnosis is the optimal solution to prevent tumor progression and allow a successful medical intervention, save life and reduce cost. This calls for early detection and analysis of breast cancer so as to increase the probability of survival and decrease the mortality rate.

## 2. Related Work

The prediction of cancer at early stages is essential as it can simplify the subsequent clinical requirements of patients and determining the effective treatment [12]. Various techniques such as clinical (traditional) techniques [21] and machine learning algorithms [22] have been proposed by researchers for early detection and prediction of breast cancer.

### 2.1 Conventional Breast Cancer Diagnosis Techniques

The traditional techniques such as Magnetic Resonance Imaging (MRI), Biospy and ultra-sound have been employed by different researchers in cancer prediction. This is particularly in determining the tumor behavior and whether a tissue is malignant or benign. Here, benign is a non-invasive types of tumor which rarely causes life threatening issues in contrast with malignant tumor which is invasive kind and is considered as a life threatening issue [23]. However, these traditional breast cancer detection techniques suffer from various setbacks. For instance, biopsy techniques prove to be painful to patients [14], while mammogram suffers from noise and is not reliable for detection of breast cancer. As discussed by Moloney et al. [24], mammography to digital breast cancer exploit attenuation of X-rays as they pass through breast tissue and it remains a good standard investigation of symptomatic women aged 40 years. Although mammography is cost effective, it has some limitations such as discomfort breast compression and radiation exposure. On the other hand, a study by Sharma et al. [25] showed that MRI achieves 96.2% sensitivity and 83.3% for clinical response and pathological respectively. However, MRI is not effective for breast cancer detection as specificity to lesion characteristics remains low, rendering distinction between cancer and benign pathology a challenge [26]. Regarding Microwave technology, Moloney et al. [24] explain that it is a cost effective, offers potential non-ionizing high resolution and is deemed to offer additional adjunct to mammography. However, this technique proves to be ambiguous in detecting of cancerous lesions [27]. Further, traditional techniques in cancer diagnosis are prone to errors and consumes more time in data processing [14,28]. Table 1 shows a summary of clinical breast cancer diagnostic techniques.

Since the traditional techniques for early cancer detection prognosis fall short in one way or another, machine learning algorithms have been proposed by various researchers to overcome challenges inherent in the traditional cancer prediction techniques.

**Table 1.** Clinical Breast Cancer Diagnosis Techniques

| Author | Techniques | Strengths | Challenges |
|---|---|---|---|
| Maloney et al. [24] | Microwave Image Technology | Cost effective<br>Offer additional adjunct to mammography<br>Potential non-ionizing, non-compressive approach | Ambiguous in detecting of cancerous lesions |
| Maloney et al. [24]<br>Zerouaoui & Idri [27] | Mammography | Cost effective | Radiation exposure<br>Uncomfortable breast compression<br>Low sensitivity |
| Maloney et al. [24]<br>Yu et al. [28]<br>Zerouaoui & Idri [27] | MRI | Has no radiation exposure<br>Provide excellent images with high contrast and resolution under appropriate condition | Low specificity to lesion characteristic rendering<br>A challenge in rendering distinction between cancer and benign pathology |
| Zerouaoui & Idri [27] | Ultrasound | Painless<br>Patients aren't exposed to ionizing radiation | Expensive |
| | Thermography | Large areas can be scanned fast<br>Not costly<br>Painless | Difficult to obtain accurate data |

## 2.2 Machine Learning Algorithms for Breast Cancer Diagnosis

Machine learning is the science of designing and applying algorithms that can learn and work on any activity from past experience [13,29]. These algorithms have been applied in diagnosis of several cancer diseases such as leukemia, prostate, cervical and breast cancer [30]. Previous studies have utilized machine learning using binary classification for detection of certain cancer such as lung cancer, brain, stomach, skin, kidney and breast cancer [18,31]. Recently, various machine learning, Artificial Intelligence (AI) and neural schemes have been adopted for image processing. These include supervised and unsupervised machine learning techniques such as Support Vector Machine (SVM), Convolutional Neural Network (CNN), K-means, K-Nearest Neighbor (KNN), Fuzzy C-Means, Random Forest (RF), Naïve Bayes (NB), and Logistic Regression (LR). The growing popularity of machine learning algorithms stem from the fact that they play critical roles in disease diagnostics and prediction [32]. They aid in early diagnosis of diseases which not only increases the survival chances but can also control the diffusion of cancerous cells in the body. In so doing, they help patients in making decision on treatment options, hence improving overall quality of lives [33]. Figure 1 shows the classification of machine learning algorithms that have been deployed in breast cancer diagnostics.

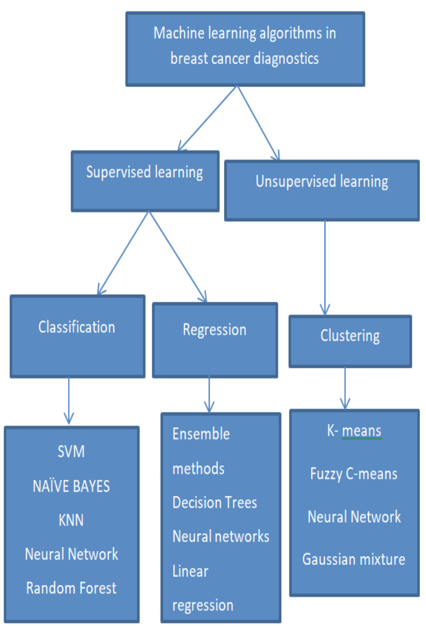


**Figure 1.** Classification of machine learning algorithms in breast cancer diagnostics

Jasti et al. [12] have proposed Least Square-Support Vector Machine, K-Nearest Neighbor, Random Forest and Naïve Bayes to aid in classification and identification of breast cancer. This model combines image pre-processing, feature extraction, feature selection and machine learning models to aid in the classification and identification of breast cancer. Here, image pre-processing is accomplished using geometric mean filter. This is followed by image enhancement and segmentation to categorize and detect breast cancer. They employed K-means clustering in order to speed-up resilient features selection. The noise was removed using image filtering technique and a geometric mean filter was used to remove noise from the input images. On the other hand, both LS-SVM and SVM algorithms help in classification and regression problems and as such, they act as non-probabilistic binary linear classifiers as they build a hyper plane which separates two classes. As explained by Michael et al. [34], LS-SVM has been used to solve linear equations and to find the training model for classification whereas SVM has been used to solve quadratic equations. However, LS- SVM proved to be less costly than SVM because it only needs to solve linear equations. In addition, LS-SVM works out well with linear-multivariate classifiers. On the other hand, deep learning based classification of breast tissues from histology images has a low accuracy. This is attributed to lack of training data and lack of knowledge on structural and textual data. However, SVM, LS-SVM, NB KNN and RF are expensive in terms of processing time [12].

On the other hand, Michael et al. [34] employed LS-SVM and SVM to solve linear equations and quadratic equations respectively. Similarly, Kumar et al. [35] have proposed KNN, Logistic Regression (LR) and SVM to differentiate between benign and malignant tumors. The authors employed KNN, logic regression and support vector models to differentiate benign and malignant tumors. In their study, they employed Wisconsin breast cancer data set for diagnostic from Kagle website. The data set contained 699 instances from different patients. They used classification technique to differentiate benign from malignant tumors. 5-fold cross-validation technique was employed in all models on training data using MATLAB.

In addition, Ara et al. [3] have employed Relief algorithm and Auto-encoder PCA algorithm for breast cancer prediction where they employed K-fold cross validation [36] for model validation and best Hyper parameter selection. The results obtained showed that SVM model performed better in terms of accuracy when compared to other states of the art approaches. A similar model was employed by Haq et al. [37] for breast cancer detection where Relief and Auto-encoder PCA algorithms were utilized for related feature selection training and testing of classification. The results showed that the features selected by Relief algorithm were more relevant to accurate detection of breast cancer than features selected by Auto-encoder and Principal Component Analysis (PCA).

On their part, Vy et al. [38] have utilized machine learning classification model to differentiate carcinoma in Situ and minimally invasive breast cancer. However, Karatza et al. [39] have proposed RF, NN [40] and Ensembles of Neural Network (ENN) for optimization of breast cancer diagnosis.

The KNN classifier is known as "lazy" algorithm since it does not take into account any assumptions on how data is being spread out. It figures out what the unknown pattern looks like based on its closest neighbor. On its part, the Euclidean distance is used to classify objects and measure the distance between them. However, this algorithm has a shortfall in terms of accuracy in breast cancer diagnostics due to incorrect prediction of true negative and false negative matrices [14].

Karatza et al. [39] have proposed Random Forest (RF), Neural Network (NN), and Ensembles of Neural Networks (ENN) for optimization of breast cancer diagnosis. This involves the use of interpretability methods such as Global Surrogate, Shapley values and Individual Conditional Expectation(ICE). The authors used Wisconsin Diagnostic Breast Cancer (WDBC) dataset of the open UCI repository for training and evaluation of these AI algorithms. Table 2 shows the summary of the current breast cancer diagnostic machine learning algorithms.

**Table 2.** Summary of Breast Cancer Diagnosis Machine Learning Algorithms

| Author | Algorithm | Strengths | Techniques |
|---|---|---|---|
| Haq et al. [37] | SVM | Effective in high dimensional space<br>Decision function<br>Versatile<br>Effective in situations where the number of measurements exceeds the number of sample | Feature elimination<br>Extraction<br>Classification<br>K-fold cross validation |
| Jasti et al. [12]<br>Haq et al. [37] | KNN | Helps to enhance classification<br>yielded higher prediction<br>accuracy than SVM<br>Lazy learner (Instance-based<br>earning)<br>new data can be added seamlessly<br>Does not make any assumption about data spread out<br>Simple to implement | 5-fold cross validation<br>PCA |
| Haq et al. [37] | Logistic Regression | Accurate | K-fold cross validation<br>Optimized hyper-parameter |
| Vy et al. [38] Jasti et al. [12] | CNN | Adjusts weights and biasness in breast cancer diagnostics<br>Accuracy in image recognition problems | Back propagation<br>classification |
| Karatza et al. [39]<br>Haq et al. [37] | Random forest | Improved accuracy | Global surrogate |
| Karatza et al. [39] | Neural network | High performance of feature extraction | Individual conditional expectation |
| Karatza et al. [39] | Ensembles of neural network | Improved accuracy<br>Strong generalization ability | Shapley values |
| Jasti et al. [12] | K-means | Capacity to cluster numerical and non-categorical data<br>Consistent in terms of computational time | Euclidean distance |
| Jasti et al. [12] | Fuzzy C-means | Scientists can model non-linear, imprecise and complex systems | Fuzziness value<br>Termination criteria |
| Jasti et al. [12] | LS-SVM | Works out well with linear –multivariate classifier<br>Help in classification and regression problem<br>Solve linear equation | Feature extraction |
| Jasti et al. [12] | Naïve Bayes | Does not require much training data<br>Simple and easy to implement | Image-preprocessing |
| Jasti et al. [12] | Relief algorithm | Solve quadratic equation<br>Less costly | Feature selection |
| Ara et al. [3] | Auto-encorder PCA algorithm | Easy to implement<br>Computationally fast | K-fold cross validation |

## 3. Empirical Review of Conventional Machine Learning Algorithms

Various researchers have employed a number of performance metrics for evaluation of machine learning models. These metrics include accuracy, precision, recall and F-score. The results have shown that semi-supervised learning models have higher accuracies (90% to 98%) with half of the training data, where K-nearest Neigbour (KNN) model for Supervised Learning (SL) and logistic regression of Semi-Supervised learning (SSL) achieved the highest accuracy of 98%. Sammut et al. [41] have employed accuracy, precision and sensitivity for evaluating the performance of KNN, SVM and LR algorithms. The results show that SVM achieved the highest accuracy of 97.7% with quadratic and cubic models. The positive prediction rate for benign was 97% and 99% for malignant. On the other hand, the false prediction rate for benign was 3% and 1% for malignant. However the algorithms consumed a lot of time for processing data.

Haq et al. [37] employed accuracy, specificity, sensitivity and F1 measure performance metrics for evaluating SVM and Relief SVM models in breast cancer diagnostics. SVM linear performance/full feature achieved 97.22% accuracy, 95% specificity, 89% sensitivity, 97% F1 measure and 0.0037 second processing time. SVM linear with selected features obtained 99.91% accuracy, 99% specificity, 100% sensitivity, 99% F1-measure and 0.02 second for processing. Relief-SVM achieved high accuracy of 99.91%, which is due to appropriate feature selection of Relief (Forest selection) algorithm and SVM predictive model. On the other hand, Vy et al. [38] evaluated their model using performance metrics such as sensitivity, specificity, accuracy, precision and F1 score. The developed model performed well reaching the accuracy of 0.84 (95% confidence interval 0.76-0.91), operating characteristic curve (AUC) of 0.93 (95% confidence interval 0.87-0.95), specificity of 0.75 (95% confidence interval 0.67-0.83) and sensitivity of 0.91 (95% confidence interval 0.76-0.94).

On their part, Karatza et al. [39] employed accuracy, sensitivity, specificity and Area under Curve (AUC) to evaluate RF, NN and ENN models. According to their study, ENN achieved 96.6% accuracy and 0.96 Area under Receiver Operating Characteristic (ROC) curve. Feature selection based on feature's importance according to Global Surrogate (GS) model improved accuracy performance of RF from 96.49% to 97.18% and area under ROC curve from 0.96% to 0.97%. On the other hand, feature selection based on feature's importance according to SVM improved the accuracy performance of NN from 94.6% to 95.53% and the area under ROC curve from 0.94 to 0.95%. Further Karatza et al. [39] noted that RF achieved 96.49% accuracy, 94.37% sensitivity, 97.75% specificity and AUC 0.96%. On the other hand, NN attained 94.1% accuracy, 88.36% sensitivity, 98.74% specificity and 0.93% AUC. Howwver, ENN achieved 96.6% accuracy, 94.94% sensitivity, 98.48% specificity and 0.96%AUC. Sammut et al. [41] and Haq et al. [37] employed accuracy performance matrix for the evaluation of Support Vector Machine (SVM) with feature elimination models for breast cancer detection in young women. The schemes yielded high accuracy of 97.7% and 99.91% respectively. Table 3 shows performance of classical breast cancer machine learning algorithms.

**Table 3.** Performance of Classical Breast Cancer Machine Learning Algorithms

| Author | Machine learning algorithm | Attained Performance |
|---|---|---|
| Haq et al. [37] | SVM | Accuracy 97.22%<br>Sensitivity 89%<br>Specificity 95%<br>FI- measure 97%<br>Time 0.0037 seconds |
| Haq et al. [37] | SVM linear with selected feature | Accuracy 99.9%<br>Sensitivity 100%<br>Specificity 99%<br>FI- measure 97%<br>Time 0.02 seconds |

Table 3 continued

| Author | Machine learning algorithm | Attained Performance |
|---|---|---|
| Vy et al. [38] | Machine learning classification | Accuracy of 0.84 (95% confidence interval 0.76-0.91)<br>Specificity of 0.75 (95% confidence interval 0.67-0.83)<br>sensitivity of 0.91 (95% confidence interval 0.76-0.94)<br>AUC of 0.93 (95% confidence interval 0.87-0.95) |
| Karatza et al. [39] | Random Forest | Accuracy 96.49%<br>Sensitivity 94.37%<br>Specificity 97.75%<br>AUC 0.96% |
| | Neural network | Accuracy 94.4%<br>Sensitivity 88.36%<br>Specificity 98.74%<br>AUC 0.93% |
| | Ensembles of Neural Network | Accuracy 96.6%<br>Sensitivity 94.94%<br>Specificity 98.48%<br>AUC 0.96% |
| Sammut et al. [41] | SVM | Accuracy 97.7% |
| | Feature elimination | Accuracy 99.91% |
| Kumar et al. [35] | Logistic regression | Accuracy 94.4% |
| | KNN | Accuracy 95.8% |
| | Decision Tree | Accuracy 95.1% |
| | Naïve Bayes | Accuracy 92.3% |
| | Random Forest | Accuracy 96.5% |
| | SVM | Accuracy 96.5% |

## 4. Challenges with Conventional Machine Learning Algorithms

Although various researchers have proposed various diagnostic models and algorithms for detection and prediction of breast cancer, these models still suffer from some setbacks specific to their implementations. As such, they need some improvement for the efficient prediction of breast cancer at early stages. For instance, the current machine learning algorithms exhibit low prediction accuracies and elongated execution time. It has been observed that K-means algorithm employs Euclidean distance scheme to determine the distance between cluster centers and data points. However this scheme does not guarantee high accuracy when executed in different iterations. In C-means algorithm, clusters are identified on similarity basis in which each data points belongs to a single cluster. Fuzzy c-means algorithm employs fuzziness value and termination criteria to determine the execution time. However, these schemes are extensive in terms of computational time due to several iterations and fuzzy measures calculations. Convolutional Neural Network (CNN) employs back propagation scheme in adjusting weight and biases in breast cancer diagnostics. However the back-propagation scheme is slow due to frequent retraining of data. The SVM with feature elimination technique for evaluation of non-occurrence and re-occurrence of breast cancer in young women is shown to be expensive in terms of execution time. On the other hand, the KNN with 10 fold cross validation is expensive in terms of processing time as the testing phase demonstrated is slow and difficult in predicting the new data. This is due to the fact that the algorithm finds only the nearest neighbor from training data. Table 4 shows challenges with legacy breast cancer machine learning algorithms.

**Table 4.** Challenges with Legacy Breast Cancer Machine Learning Algorithms

| Author | Machine learning algorithm | Weaknesses |
|---|---|---|
| Sammut et al. [41]<br>Haq et al. [37] | SVM | Consumes time<br>Difficult to determine optimal parameters<br>Do not provide projections of probability<br>straight away |
| Jasti et al. [12]<br>Sammut et al. [41]<br>Haq et al. [37] | KNN | Slow<br>Difficult in predicting data<br>High risk with larger data set<br>High dimension need feature scaling<br>Sensitive to noise data, missing values and outliers |
| Sammut et al. [41]<br>Haq et al. [37] | Logistic Regression | High computation time |
| Vy et al. [38]<br>Jasti et al. [12] | CNN | Slow due to frequent retraining<br>High computational cost<br>Need a lot of training data |
| Karatza et al. [39] Sammut et al. [41]<br>Haq et al. [37] | Random Forest | High computation time |
| Karatza et al. [39] | Neural network | Time complexity<br>Lack of hybridized classifiers |
| Karatza et al. [39] | Ensembles neural network | Time complexity<br>Unable to effectively compute high dimensional task |
| Jasti et al. [12] | K-means | Does not guarantee high accuracy when executed in different iteration |
| Jasti et al. [12] | Fuzzy C- means | Expensive in terms of computational time |
| Jasti et al. [12] | LS-SVM | High execution time |
| Jasti et al. [12] | Naïve Bayes | High execution time<br>Redundant attributes which misleads classification |
| Jasti et al. [12] | Relief algorithm | High execution time |

It is evident that machine learning algorithms suffer from many limitations in terms of accuracy and time complexity. To overcome these challenges, new approaches are required for efficient and accurate detection and prediction of breast cancer with minimal execution time.

## 5. Current Trends in Machine Learning Based Cancer Predictions

It is evident that most of the conventional classical machines learning algorithms face various challenges during breast cancer diagnosis. Quantum Machine learning (QML) algorithms can potentially help alleviate some of these issues such as time complexity [42] and storage capacity through the exploration of quantum computing properties such as entanglement and superposition. Quantum can be described as the amount, quantity or a portion and as such, the world of atoms and elementary particles is characterized by the phenomena that energy can only be released or absorbed in a defined amounts “quanta” [43]. On the other hand, quantum computing is regarded as an area in computer science that employs the principles of quantum theory, which explains the nature and behavior of energy and material on the atomic and sub atomic levels [39]. To describe the behavior of quantum particles, physicist created a theory of quantum mechanics known as “quantum theory”. Based on this theory, the potential energy determines the particle distribution [12].

According to Azevedo et al. [44], quantum computing is a new research area which employs quantum mechanics such as superposition, entanglement and interference concepts. This quantum approach makes certain kinds of problems to be solved easier as compared to classical computing such as machine learning algorithms. This is due to its operation on quantum gates (qubits) as compared to its counterpart (classical computers) which employs traditional bits. Quantum computing has been effectively employed in various domains such as classification, prediction, object detection and tracking. These quantum-based techniques have remarkable performance over classical theory based models. As such, quantum computing can be used to solve many existing problems in different fields such as machine learning deployed for security, biomedical (diabetes), financial markets and weather forecasting. This is due to parallel computing inherent in them and the utilization of quantum computations in solving machine learning tasks [45].

Various past research works have shown strong evidence that quantum machine learning algorithms could outperform classical algorithms in solving certain machine learning problems [12]. Examples of such algorithms include Quantum Support Vector Machine (QSVM), Quantum Convolutional Neural Network (QCNN) and Quantum Principal Component Analysis (QPCA) Quantum Enhanced Feature Space (QEFS) and Quantum Generative Models (QGM). These algorithms have been employed in disease diagnostics and have made remarkable improvements. For instance in QSVM, higher dimension vector space optimization boundary is used to classify the classes of labeled data and PCA algorithm. This makes a pattern of huge unlabeled data and effectively reduces it, making it easier for further analysis. QPCA is unsupervised model and it's also a mathematical technique that transforms a set of uncorrelated variables principal components. It is also used as a feature extraction technique that reduces the dimensionality of the database and to build an effective quantum machine learning model.

The major advantage of quantum computing is the increasing number of qubits of a quantum computer [46]. As such, the process power increases exponentially. Since machine learning algorithms have shown to be powerful tools for finding patterns in data, quantum systems have also demonstrated to be effective in finding patterns that classical systems or algorithms are thought not to produce efficiently. Although, machine learning algorithms [47] have been employed to diagnose breast cancer with great accuracy however no studies have shown ML algorithms with improved speed in processing data. Quantum computing has been shown to work simultaneously, considering the possible solutions to a problem all at one time [48]. On the other hand, classical algorithms perform their tasks sequentially, one after the other [49]. This requires long computing times, rendering their operations time consuming. This is considered advantageous to quantum computing as it will enable solutions to very complex optimization problems. In addition, quantum gates are reversible as compared to classical algorithms. Therefore quantum computing in combination with machine learning could in principle find solutions more quickly. These models can hence be used for tasks that classical algorithms would not be able to solve in a reasonable amount of time.

Quantum supremacy becomes unprecedentedly urgent and important to explore the utilization of quantum computations in solving machine learning tasks [44]. Due to parallelism in their design, quantum algorithms are able to solve problems inherent in classical algorithms, such as speed and space. As Martina et al. [50] explain, quantum computations exhibits promising applications in machine learning and data analysis with much more advance in time and space complexity [51]. The entanglement property in quantum computing provides a mechanism for improving the storage capacity as well as retrieving corrupted or incomplete information [52]. These properties offer significant speed up of any computation evaluated on the basis of complexity. Some of the promising applications of quantum computing include health (diabetes) prediction, breast cancer, fraud detection, optimization in improving performance and prediction [43].

Previous research works have employed Quantum Machine (QM) to enhance Machine Learning (ML) and Deep Learning (DL) models performance. For instance, Gupta et al. [46] employed QML for prediction of diabetes in PIMA Indian Diabetes Dataset (PIDD). On the other hand, quantum inspired evolutionary algorithm with binary real representation has been implemented to improve ANN by providing self-configuring ability on data. Similarly, quantum nearest mean classifier has utilized the binary classification where quantum version is employed to estimate the optimal performance measures [43]. Early studies in the area of QML by Zhang and Ni [22] have shown that QML is one of the most promising technologies of research in the future. The experimental demonstration in the field of quantum computing has made remarkable interests in the recent years, giving rise to new striking possibilities of enhancing machine learning with quantum devices [53]. Examples of these models include quantum principle component analysis, quantum enhanced feature space, quantum generative models, Quantum Convolutional Network and quantum support vector machines.

Since quantum machine learning is an interdisciplinary field that explores the properties of machine learning and quantum physics, quantum mechanics have been used effectively in various domains such as prediction, object detection, classification and tracking. These models have achieved remarkable performance over classical theory based models [48]. This has been necessitated by large computing corporations such as IBM, Google and Microsoft as well as small start-ups such as Dwave, Regetti and IonQ which have made quantum computing easier. This is due to provision of access through the cloud free of charge.

In the era of big data, Machine Learning (ML) may benefit from either speed, complexity or smaller amounts of storage through the exploration of quantum computing properties such as entanglement and superposition [44]. As such, quantum computing has recently been incorporated in ML algorithms such as Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN). For instance, QML regards CNN as the best algorithm with regard to image content identification and have exhibited exemplary performance in several tasks. However it suffers from high complexity. This is particularly the case when recognizing more complex features as they end up aggregating and recombining. Therefore, improving the speed of networks can have a huge impact when training [54] models that require detailed images as input like mammogram [52,55,56].

## 6. Proposed Quantum Machine Learning Based Breast Cancer Diagnostic Algorithm

The prediction accuracy for breast cancer diagnostic model requires more enhancements for efficient and accurate detection at early stages. This will facilitate better treatment and recovery. As such, this paper recommends a quantum based machine learning algorithm which will be more efficient in terms of prediction accuracy with minimal execution time. This is inspired by recent advances in physical quantum processor.

In the proposed QML algorithm, a number of steps are executed during the breast cancer data classification using quantum machine learning classifier. These steps include the imputing of breast cancer dataset after which data processing is done. This is followed by the feature reduction procedures after which the reduced data is partitioned into training and testing datasets. Next, data translation is carried out, followed by the invocation of the QML classifier that is trained using the quantum system. Afterwards, the hyper-parameters are optimized. Finally, the performance of the developed classifier is evaluated. Figure 2 presents the summary of these procedures.

The detailed descriptions of these steps are illustrated below.

*Input breast cancer dataset:*Quantum machine learning is data-driven and therefore the first step in building an efficient classifier involves the feeding of adequate amount of data. This breast cancer falls into two categories, which include benign and malignant.

*Data pre-processing:* Pre-processing refers to the initial procedures that are meant to make the data suitable for subsequent model building. In this step, issues such as missing data in the dataset are handled. It may also involve expunging of irrelevant fields in the underlying dataset. Afterwards, scaling and normalization are carried out to transform the feature space into relatively the same range.

*Feature reduction:* In the proposed algorithm, the training and testing is executed using a quantum system which has a restricted number of qubits. It is therefore required that the dimensionality of the original dataset be minimized. Here, the Principal Component Analysis (PCA) algorithm shall be deployed for feature reduction. The choice of the PCA is due to its dimensionality reduction which makes it easier to explore and visualize data and it can solve in tracking the problem of the limited number of qubits. On its part, the algorithm shall identify patterns in data and highlight their similarities and differences and at the same time it can compress data without much loss of information. It can efficiently eliminate irrelevant influencing factors.

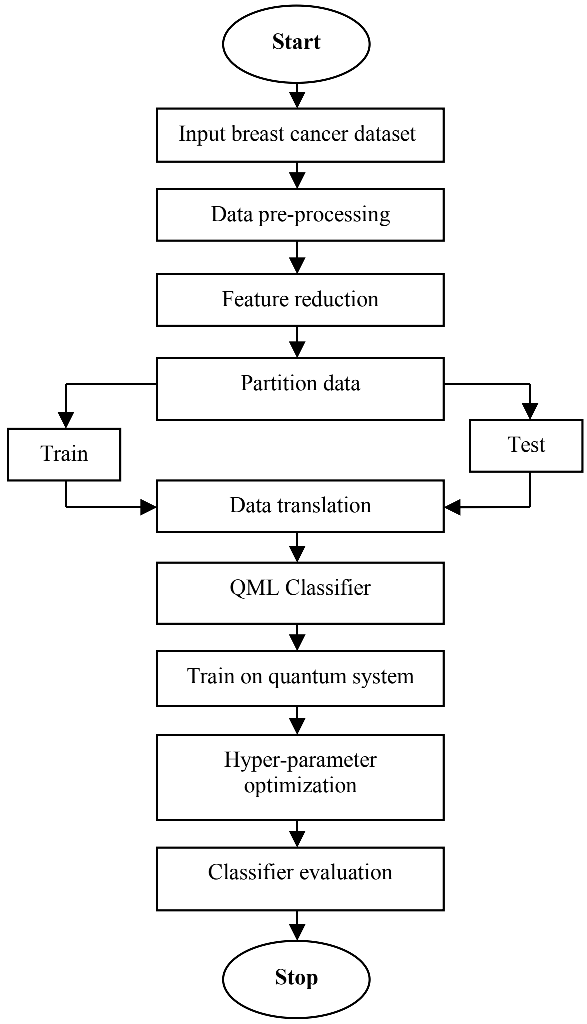


**Figure 2.** Proposed Breast Cancer Diagnostic framework

*Data Partitioning:* In this step, the minimized and transformed data serves as the final dataset that is then utilized for training and testing.

*Data translation:*In this step, a data point such as α is transformed into quantum data point  using a circuit C (). Here, λ () represents any classification function applied to data point . Afterwards, a parameterized quantum circuit P (ω) with ω parameters is utilized to process these data items. Next, the respective labels for the transformed data are obtained by the application of measurements that output classical value of -1, 1 for every classical input .

*Training:* to train the QML classifier, 10-fold cross validation is applied so as to yield a refined and fair performance metrics. The process of training QML involve obtaining the quantum kernel and building the feature map.

*Hyper-parameter optimization:* to fine tune the breast cancer diagnosis process, the depth and feature dimensions need to be optimized. This optimization will be accomplished using the inbuilt features of the platform that will be utilized to execute this algorithm. This will basically involve determining the best configuration that will make the diagnostic model to yield minimal loss.

## 7. Conclusions and Future Research Directions

Machine learning algorithms have been developed by various researchers in disease prognosis and more specifically in breast cancer prediction and diagnosis. These algorithms have proved to be of significance and more specifically in early breast cancer diagnosis. They have aided in minimizing human errors and delivered prompt analysis of medical data with greater depth. In essence, classifier based prediction models riding on data mining and ML can limit the diagnosis errors and enhance the efficiency of a cancer diagnosis. However these algorithms have fallen short in terms of prediction accuracy due to incorrect predictions. This is reflected in their erroneous true negatives, false negatives and long execution times. QML algorithms will overcome many challenges of classical machine learning. Experimental demonstrations in the field of quantum computing have also made remarkable contributions in recent years, giving rise to new striking possibilities of enhancing machine learning with quantum devices. As such, machine learning may benefit from speed, complexity and smaller amounts of storage through the exploration of quantum computing properties such as entanglement and superposition. Based on these strengths, this paper has proposed the development of quantum machine learning algorithm for breast cancer diagnostics. This new approach integrates the principles of quantum computing with legacy machine learning algorithms. Future work will involve the practical realization of the proposed algorithm for breast cancer diagnosis, during which its performance will be evaluated. Thereafter, comparisons will be carried out against conventional breast cancer prediction techniques and algorithms.

## Conflict of Interest

There is no conflict of interest.

## References

[1] Al-Azzam, N., Shatnawi, I., 2021. Comparing supervised and semi-supervised machine learning models on diagnosing breast cancer. Annals of Medicine and Surgery. 62, 53-64.

[2] Rahman, M.M., Ghasemi, Y., Suley, E., et al., 2021. Machine learning based computer aided diagnosis of breast cancer utilizing anthropometric and clinical features. Irbm. 42(4), 215-226.

[3] Ara, S., Das, A., Dey, A., 2021, April. Malignant and benign breast cancer classification using machine learning algorithms. 2021 International Conference on Artificial Intelligence (ICAI). pp. 97-101. IEEE.

[4] Thalor, A., Joon, H.K., Singh, G., et al., 2022. Machine learning assisted analysis of breast cancer gene expression profiles reveals novel potential prognostic biomarkers for triple-negative breast cancer. Computational and Structural Biotechnology Journal. 20, 1618-1631.

[5] Safdar, S., Rizwan, M., Gadekallu, T.R., et al., 2022. Bio-Imaging-Based Machine Learning Algorithm for Breast Cancer Detection. Diagnostics. 12(5), 1134.

[6] Nyangaresi, V.O., El-Omari, N.K.T., Nyakina, J.N., 2022. Efficient Feature Selection and ML Algorithm for Accurate Diagnostics. Journal of Computer Science Research. 4(1), 10-19.

[7] Chugh, G., Kumar, S., Singh, N., 2021. Survey on machine learning and deep learning applications in breast cancer diagnosis. Cognitive Computation. 13(6), 1451-1470.

[8] Rane, N., Sunny, J., Kanade, R., et al., 2020. Breast cancer classification and prediction using machine learning. International Journal of Engineering Research and Technology. 9(2), 576-580.

[9] Alaa, A.M., Gurdasani, D., Harris, A.L., et al., 2021. Machine learning to guide the use of adjuvant therapies for breast cancer. Nature Machine Intelligence. 3(8), 716-726.

[10] Mohammed, S.A., Darrab, S., Noaman, S.A., et al., 2020. Analysis of breast cancer detection using different machine learning techniques. International Conference on Data Mining and Big Data. pp. 108-117. Springer, Singapore.

[11] Shravya, C., Pravalika, K., Subhani, S., 2019. Prediction of breast cancer using supervised machine learning techniques. International Journal of Innovative Technology and Exploring Engineering (IJITEE). 8(6), 1106-1110.

[12] Jasti, V., Zamani, A.S., Arumugam, K., et al., 2022. Computational technique based on machine learning

and image processing for medical image analysis of breast cancer diagnosis. Security and Communication Networks.

[13] Srinivas, T., Karigiri Madhusudhan, A.K., Arockia Dhanraj, J., et al., 2022. Novel Based Ensemble Machine Learning Classifiers for Detecting Breast Cancer. Mathematical Problems in Engineering.

[14] Rasool, A., Bunterngchit, C., Tiejian, L., et al., 2022. Improved machine learning-based predictive models for breast cancer diagnosis. International Journal of Environmental Research and Public Health. 19(6), 3211.

[15] Ahuja, A., Al-Zogbi, L., Krieger, A., 2021. Application of noise-reduction techniques to machine learning algorithms for breast cancer tumor identification. Computers in Biology and Medicine. 135, 104576.

[16] Binder, A., Bockmayr, M., Hägele, M., et al., 2021. Morphological and molecular breast cancer profiling through explainable machine learning. Nature Machine Intelligence. 3(4), 355-366.

[17] Sung, H., Ferlay, J., Siegel, R.L., et al., 2021. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer Journal for Clinicians. 71(3), 209-249.

[18] Tariq Jamal, A., Ben Ishak, A., Abdel-Khalek, S., 2021. Tumor edge detection in mammography images using quantum and machine learning approaches. Neural Computing and Applications. 33(13), 7773-7784.

[19] Chaudhury, S., Krishna, A.N., Gupta, S., et al., 2022. Effective image processing and segmentation-based machine learning techniques for diagnosis of breast cancer. Computational and Mathematical Methods in Medicine.

[20] Nyangaresi, V.O., Rodrigues, A.J., Abeka, S.O., 2020, December. ANN-FL secure handover protocol for 5G and beyond networks. International Conference on e-Infrastructure and e-Services for Developing Countries. pp. 99-118. Springer, Cham.

[21] Zhang, Y., Ni, Q., 2020. Recent advances in quantum machine learning. Quantum Engineering. 2(1), e34.

[22] Nyangaresi, V.O., Ahmad, M., Alkhayyat, A., et al., 2022. Artificial neural network and symmetric key cryptography based verification protocol for 5G enabled Internet of Things. Expert Systems. pp. e13126.

[23] Houssein, E.H., Emam, M.M., Ali, A.A., et al., 2021. Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review. Expert Systems with Applications. 167, 114161.

[24] Moloney, B.M., O'Loughlin, D., Abd Elwahab, S., et al., 2020. Breast cancer detection—A synopsis of conventional modalities and the potential role of microwave imaging. Diagnostics. 10(2), 103.

[25] Sharma, U., Jagannathan, N.R., 2022. Magnetic Resonance Imaging (MRI) and MR Spectroscopic Methods in Understanding Breast Cancer Biology and Metabolism. Metabolites. 12(4), 295.

[26] Haque, M.N., Tazin, T., Khan, M.M., et al., 2022. Predicting Characteristics Associated with Breast Cancer Survival Using Multiple Machine Learning Approaches. Computational and Mathematical Methods in Medicine.

[27] Zerouaoui, H., Idri, A., 2021. Reviewing machine learning and image processing based decision-making systems for breast cancer imaging. Journal of Medical Systems. 45(1), 1-20.

[28] Yu, Y., He, Z., Ouyang, J., et al., 2021. Magnetic resonance imaging radiomics predicts preoperative axillary lymph node metastasis to support surgical decisions and is associated with tumor microenvironment in invasive breast cancer: A machine learning, multicenter study. EBioMedicine. 69, 103460.

[29] Nyangaresi, V.O., Rodrigues, A.J., 2022. Efficient handover protocol for 5G and beyond networks. Computers & Security. 113, 102546.

[30] Alzu'bi, A., Najadat, H., Doulat, W., et al., 2021. Predicting the recurrence of breast cancer using machine learning algorithms. Multimedia Tools and Applications. 80(9), 13787-13800.

[31] Li, J., Zhou, Z., Dong, J., et al., 2021. Predicting breast cancer 5-year survival using machine learning: A systematic review. PloS one. 16(4), e0250370.

[32] Mirsadeghi, L., Haji Hosseini, R., Banaei-Moghaddam, A.M., et al., 2021. EARN: an ensemble machine learning algorithm to predict driver genes in metastatic breast cancer. BMC Medical Genomics. 14(1), 1-19.

[33] Meenalochini, G., Ramkumar, S., 2021. Survey of machine learning algorithms for breast cancer detection using mammogram images. Materials Today: Proceedings. 37, 2738-2743.

[34] Michael, E., Ma, H., Li, H., et al., 2022. An optimized framework for breast cancer classification using machine learning. BioMed Research International.

[35] Kumar, V., Mishra, B.K., Mazzara, M., et al., 2020. Prediction of malignant and benign breast cancer: A data mining approach in healthcare applications. Advances in Data Science and Management. pp. 435-

442. Springer, Singapore.
[36] Al Sibahee, M.A., Ma, J., Nyangaresi, V.O., et al., 2022, June. Efficient Extreme Gradient Boosting Based Algorithm for QoS Optimization in Inter-Radio Access Technology Handoffs. 2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA). pp. 1-6. IEEE.
[37] Haq, A.U., Li, J.P., Saboor, A., et al., 2021. Detection of breast cancer through clinical data using supervised and unsupervised feature selection techniques. IEEE Access. 9, 22090-22105.
[38] Vy, V.P.T., Yao, M.M.S., Khanh Le, N.Q., et al., 2022. Machine Learning Algorithm for Distinguishing Ductal Carcinoma in Situ from Invasive Breast Cancer. Cancers. 14(10), 2437.
[39] Karatza, P., Dalakleidi, K., Athanasiou, M., et al., 2021, November. Interpretability methods of machine learning algorithms with applications in breast cancer diagnosis. 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). pp. 2310-2313. IEEE.
[40] Nyangaresi, V.O., Rodrigues, A.J., Abeka, S.O., 2022. Machine Learning Protocol for Secure 5G Handovers. International Journal of Wireless Information Networks. 29(1), 14-35.
[41] Sammut, S.J., Crispin-Ortuzar, M., Chin, S.F., et al., 2022. Multi-omic machine learning predictor of breast cancer therapy response. Nature. 601(7894), 623-629.
[42] Nyangaresi, V.O., Rodrigues, A.J., Abeka, S.O., 2020, October. Neuro-fuzzy based handover authentication protocol for ultra dense 5G networks. 2020 2nd Global Power, Energy and Communication Conference (GPECOM). pp. 339-344. IEEE.
[43] Khan, T.M., Robles-Kelly, A., 2020. Machine learning: Quantum vs classical. IEEE Access. 8, 219275-219294.
[44] Azevedo, V., Silva, C., Dutra, I., 2022. Quantum transfer learning for breast cancer detection. Quantum Machine Intelligence. 4(1), 1-14.
[45] Salehi, T., Zomorodi, M., Plawiak, P., et al., 2022. An optimizing method for performance and resource utilization in quantum machine learning circuits. Scientific Reports. 12(1), 1-16.
[46] Gupta, H., Varshney, H., Sharma, T.K., et al., 2022. Comparative performance analysis of quantum machine learning with deep learning for diabetes prediction. Complex & Intelligent Systems. 8(4), 3073-3087.
[47] Nyangaresi, V.O., Rodrigues, A.J., Abeka, S.O., 2020. Secure Handover Protocol for High Speed 5GNetworks. International Journal of Advanced Networking and Applications. 11(06), 4429-4442.
[48] Li, W., Lu, S., Deng, D.L., 2021. Quantum federated learning through blind quantum computing. Science China Physics, Mechanics & Astronomy. 64(10), 1-8.
[49] Sharifi, M., Hasan, A., Attar, F., et al., 2020. Development of point-of-care nanobiosensors for breast cancers diagnosis. Talanta. 217, 121091.
[50] Martina, S., Buffoni, L., Gherardini, S., et al., 2022. Learning the noise fingerprint of quantum devices. Quantum Machine Intelligence. 4(1), 1-12.
[51] Nyangaresi, V.O., Abeka, S.O., Rodgrigues, A., 2018. Secure Timing Advance Based Context-Aware Handover Protocol for Vehicular Ad-Hoc Heterogeneous Networks. International Journal of Cyber-Security and Digital Forensics. 7(3), 256-275.
[52] Li, R.Y., Gujja, S., Bajaj, S.R., et al., 2021. Quantum processor-inspired machine learning in the biomedical sciences. Patterns. 2(6), 100246.
[53] Pomarico, D., Fanizzi, A., Amoroso, N., et al., 2021. A proposal of quantum-inspired machine learning for medical purposes: An application case. Mathematics. 9(4), 410.
[54] Nyangaresi, V.O., Abeka, S.O., Rodrigues, A.J., 2020. Delay Sensitive Protocol for High Availability LTE Handovers. American Journal of Networks and Communications. 9(1), 1-10.
[55] Yala, A., Mikhael, P.G., Strand, F., et al., 2022. Multi-institutional validation of a mammography-based breast cancer risk model. Journal of Clinical Oncology. 40(16), 1732-1740.
[56] Chakravarthy, S.S., Rajaguru, H., 2022. Automatic detection and classification of mammograms using improved extreme learning machine with deep learning. IRBM. 43(1), 49-61.