International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 October 2023*

Breast Cancer with Deep Learning Using Feature Selection: A Systematic Literature Review

Charusharma[1*] and Dr. Kavita Gupta[2]
University Institute of Computing, Chandigarh University
University Institute of Computing, Chandigarh University

***Corresponding Author:-** Charusharma
*Email: charusharma2009@rediffmail.com

Abstract : Among the most hazardous illnesses for people is cancer, yet there is currently no long-term treatment available. One of the typical cancers is breast cancer. More than 276,000 new instances of invasive breast cancer and much more than 48,000 non-invasive instances were detected in India in 2022 alone, based on the National Breast Cancer Foundation. Considering that 64% of these instances are discovered early in the course of the illness, patients have a 99% probability of surviving. A systematic review was performed to understand various deep learning algorithms/models, classes and classification of diagnosis, accuracy rate based on those algorithms/models, databases, methods and performance evaluation parameters. Though various research articles from reputed journals have been reviewed, a final in-depth review was considered on 45 numbers of papers by eliminating the irrelevant research articles based on some filtration criteria. The results and discussions are provided which reveals the current trends and adoptions by the various researchers conducting their research on breast cancer diagnosis using deep learning feature selection technology.

INTRODUCTION

Typically, cancer is identified as the destruction of healthy cells and abnormal human cell development. In the case of breast cancer, abnormal cells infiltrate nearby cells and spread to other areas of the body. There are two forms of breast cancer: benign and malignant, also known as semi-invasive or invasive. Many deep learning researchers begin their work by determining if a tumor is malignant or not to determine the severity of breast cancer. Deep learning is a method of using little to no human interaction to make data judgments, which is a part of artificial intelligence (AI). It uses data analysis to find patterns, make judgments, and create analytical models. In terms of medical data, it is a subset of health information based on patient safety or clinical trial protocols, and includes patient electronic health records.

AI is capable of collecting data from health records, processing it, and providing a clear result. The algorithm in this process identifies patterns and provides its own reasoning. The primary goal of the AI algorithm is to find the link between prevention or treatment and the patient's diagnosis. Classification is the process of identifying which class a given instance of data belongs to. Binary classification and multiple class classification are also possible techniques. Regression is the process of using a related set of characteristics to attempt to predict a label, which is a continuous value. Clustering is the process of classifying data instances into several groups based on how similar they are. Anomalies, or rare or occasional occurrences that are misleading or different from the majority of observations, can also be identified through anomaly detection. In ranking, labeled data is organized into instances and given rankings, which are then used by the ranker to determine ranks for unseen examples. The task of suggesting items or services to a user based on their past data is called recommendation. Forecasting is the process of making predictions about the future based on historical time-series data.

Predictive modeling is a central aspect of deep learning. Predictive models are trained on historical data to make predictions for new, unseen data. The performance of the model is determined by the effectiveness of the method used to solve the problem. Deep learning algorithms perform well when given the right data. To improve the performance of these models, feature selection techniques are often used. Feature selection, also known as variable selection, is the process of selecting a subset of variables or features from a larger dataset to create deep learning models. This helps in reducing computing costs and improving the performance of the model. Simpler models require less training time and are easier to understand. By using fewer variables, the cost of computing and model construction is reduced. Additionally, feature selection promotes generalization, which reduces model overfitting. Removing irrelevant or noisy variables can also decrease the chance of data collection errors and improve generalization.

Predictive modeling forms the majority of deep learning models. To create predictions for new, unseen data, a predictive modeling algorithm is trained on past data. The effectiveness of the approach used to solve the problem determines how well the supervised neural model performs. When given the right data, deep learning algorithms perform efficiently. In this regard, feature selection methods are very useful. They not only help in reducing computing costs, but also help the model perform better. The process of choosing a set of variables or features from the complete dataset to create deep learning models is called feature selection or variable selection. It is crucial in building deep learning models that are faster, simpler, and more straightforward. Simpler models require less training time and are easier to understand. A model with ten variables is easier to understand than one with one hundred variables. Using fewer variables also lowers the cost of computing and speeds up model construction. Additionally, feature selection promotes generalization, which reduces model overfitting.

Breast cancer is a major public health concern and early detection is crucial for improved outcomes. Given the rapid advancements in medical technology, it is imperative to keep pace with the latest developments in breast cancer detection techniques. A systematic literature review would provide an in-depth examination of existing research on the subject, enabling us to identify strengths and weaknesses of current methods, as well as to identify areas in need of further research. Furthermore, by synthesizing the available evidence, we can gain a comprehensive understanding of the state of the field and make informed recommendations for future directions. In light of these considerations, we propose a systematic literature review to gain a deeper understanding of breast cancer detection techniques and to ultimately improve the accuracy of diagnoses for patients. It is important to understand the research articles carried out their research on breast cancer detection techniques in order to increase the accuracy of diagnoses. This study was motivated by the following reasons:

1- The previous researchers have adopted different screening techniques. Different types of screening techniques have different strengths and limitations. It is very much essential to make informed decisions about which techniques are best suited for detecting breast cancer. The unique characteristics of each screening techniques for the diagnosis of breast cancer is also required to be studied (Saoud et al., 2019).
2- Another driving factor is the utilization of deep learning methods for breast cancer detection. As deep learning methods are commonly utilized in medical research and the advancement of modern feature selection techniques, it is imperative to assess and compare these methods for the examination of breast cancer to determine the most suitable approach for identifying breast cancer (Kumari & Chaudhary, 2020).
3- Evaluating the effectiveness of techniques for detecting breast cancer involves using several assessment criteria, each with distinct properties. For understanding if new methods perform better or worse than older ones, it is crucial to examine the performance of different breast cancer detection evaluation methods carefully to make informed choices (Sahu & Panigrahi, 2020).

Objective:

This study focuses on performing a systematic literature review on previous literatures in the domain of breast cancer detection using deep learning feature selection technique. The main topics covered in this paper are:

(1) Theoretical aspects of breast cancer studied in the previous literatures.

(2) The various methods/strategies, risk factors, target populations, and shared datasets used in the previous studies in this domain

(3) The performance comparison of breast cancer prediction techniques based on deep learning.

(4) Potential future prospects for breast cancer detection research.

II. RELATED WORK

The study of breast cancer detection has been a topic of interest for many researchers, and several methods have been developed over time to diagnose the condition. In (Abdullah Farid, 2021), the authors reviewed various data mining and deep learning techniques for breast cancer prediction. They found that only a few studies used genetics, with most of the studies relying on imaging. The three main algorithms used in genetic breast cancer prediction were SVM, decision trees, and random forests. Meanwhile, in imaging techniques, several algorithms such as CNNs and Naive Bayes were used. The main objective of the document is to enhance the performance in cancer prognosis prediction and develop a more generalized outcome classifier for breast cancer. The document proposes a method that integrates feature selection and feature extraction methods with deep learning techniques to learn more representative characteristics from gene expression profiles. The goal is to construct a more powerful classifier for predicting clinical outcomes in breast cancer patients.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 October 2023*

The proposed approach uses two methods for feature learning: PCA (Principal Component Analysis) and autoencoder neural network.

PCA is employed as a feature selection method to reduce the dimensionality of the gene expression profiles. It performs a linear approximation of the original data and retains significant information.

After applying PCA, the resulting features are fed into an autoencoder neural network for feature extraction. The autoencoder is a nonlinear dimensionality reduction method that learns high-level and complex features by capturing non-linear associations among expressions of different genes. The autoencoder consists of an encoder and a decoder, which transform the input features into a hidden representation and then reconstruct the original features, respectively. The autoencoder is trained using the Adam optimization algorithm to minimize the reconstruction loss with a sparsity penalty.

By combining PCA and autoencoder, the proposed approach aims to learn more representative characteristics from gene expression profiles and improve the performance of the classifier for predicting clinical outcomes in breast cancer patients.The evaluation metrics used to measure the performance of the proposed method are the area under the receiver operating characteristic curve (AUC), Matthews correlation coefficient (MCC), accuracy (ACC), specificity (SP), and sensitivity (SN). These metrics are commonly used in classification tasks, especially when dealing with imbalanced datasets. The AUC measures the overall performance of the classifier, while the MCC takes into account true positives, true negatives, false positives, and false negatives to assess the quality of the predictions. ACC represents the accuracy of the classifier, SP measures the ability to correctly identify negative instances, and SN measures the ability to correctly identify positive instances. These metrics provide a comprehensive evaluation of the proposed method's performance in predicting the clinical outcomes of breast cancer patients.

The authors in (Sahu et al., 2020) focused on gene mutations to identify breast cancer. They stated that the gene prediction classification segment aims to perform gene annotation, discovery, and mutation detection to determine if cancer is present or absent. They concluded that a variety of techniques, including regression, probability models, SVMs, NNs, and deep learning, could be employed.In (Prabadevi et al., 2020), the authors present five insect-based natural inspired computing (NIC) algorithms for diagnosing diabetes and cancer, including breast cancer. They used CNN classification and private datasets in their study (Kewat et al., 2020). Based on these surveys, we can contribute to the field by learning genetic sequencing and imaging concurrently to predict breast cancer and gather further knowledge to support early detection and treatment. We can also provide guidance to researchers who want to perform studies in this field. In (Sahu et al., 2019), the authors reviewed recent studies using various imaging modalities and deep learning to tackle breast cancer. They focused on three deep learning frameworks for breast imaging modalities and attempted to provide a comprehensive overview of the research on breast cancer imaging using DLR-based CAD systems.   They found that the NIC algorithms performed well in identifying different types of cancer, with guided ABC and neural networks being combined to identify breast cancer in (Fotouhi et al., 2019). The authors emphasized that more research is needed to identify various stages of cancer and diabetes. The authors in (Haq et al., 2018) showed the usefulness of neural networks (NNs) in the categorization of cancer diagnoses, particularly in the early stages. They found that most NNs have the potential to identify malignant cells. However, the imaging technique requires significant processing power to preprocess the images. The authors in (J. Zhang et al., 2019) compared the performance of various classification techniques in the diagnosis of non-communicable diseases (NCD). The eight classification algorithms were applied to eight NCD datasets and assessed using the accuracy metric AUC. The authors found that the KNN, SVM, and NN algorithms were robust against the noise in the NCD datasets. They also claimed that using appropriate pre-processing strategies could solve the irrelevant feature problem, leading to improved accuracy. In (Asri et al., 2016), the authors present a review of different methods for classifying breast cancer using histological image evaluation. These methods are based on several artificial neural network (ANN) designs. Thermographic images have the potential to contribute to the early detection of breast cancer. When used as a screening tool, thermography can help identify abnormal temperature patterns in breast tissue. These patterns may indicate the presence of a tumor or other abnormalities. Early detection of breast cancer is crucial for successful treatment, and thermographic images can aid in the early identification of potential tumors. By detecting abnormalities at an early stage, thermography can help prevent the spread of cancer and improve the chances of positive outcomes for patients. "The absence of prognostic models makes it difficult for medical professionals to devise treatment strategies that have the potential to lengthen a patient's overall survival time. Therefore, time is required to discover the strategy that produces the least amount of error in order to improve accuracy. Because the currently available methods to identify breast cancer, such as mammograms, ultrasounds, and

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 October 2023*

biopsies, take a significant amount of time, there was a demand for a computerised diagnostic system that utilised the technique of machine learning. This methodology makes use of algorithms that speed up the process of classifying the tumour, improve the accuracy with which cells are located, and shorten the amount of time required to do so." "In recent years, thermography has become an increasingly widespread method, particularly for the detection of cervical cancer [12]. This is because of the appealing realities from its own relatively safe invention, in addition to the chance of future upgrades made possible by cutting-edge technical improvement. The current research being conducted in this area is to arrive at a tumour outcome that is more definitive and can be agreed upon by a large number of people, and which can be utilised as a recommendation for breast cancer screening.

Machine learning algorithms are utilized in the identification and recognition of breast tumors through the application of computer-aided diagnosis (CAD) methods. These algorithms, such as Convolutional Neural Networks (CNN), Support Vector Machines (SVM), and Random Forest, are trained on large datasets of mammography images.

The algorithms use pattern recognition techniques to analyze the images and locate potential tumors. They learn to distinguish between different classes of tumors, such as malignant and benign, or cancerous and non-cancerous. By analyzing the features and patterns in the images, the algorithms can make predictions about the presence and nature of tumors.

Additionally, pre-processing techniques may be applied to the mammography images before feeding them into the machine learning algorithms. These pre-processing techniques enhance the quality of the images and improve the success rate of tumor classification.

Overall, machine learning algorithms play a crucial role in automating the identification and recognition of breast tumors, enabling faster and more accurate diagnoses

The authors found that ANNs were first used for histological image analysis around 2012, with the two most commonly used algorithms being ANNs and PNNs. Most of the work on feature extraction has used textural and morphological features.

Methodology

The methodology of this research is focused on carrying out a systematic literature review on breast cancer detection using deep learning features selection. The step wise procedure which was followed in the systematic literature review includes the following:

1- Identification: A systematic literature review was carried out on the topic of breast cancer detection using deep learning with feature selection. The Scopus database was used to identify relevant articles using the search terms "deep learning" or "deep neural network" combined with "gene," "genomics," "microarray," "DNA," "X-ray," "mammography," "MRI," and "ultrasound." A total of 738 articles were initially identified.
2- Screening: The titles and abstracts of the 738 articles were screened to determine the relevance of each study to the research question and to remove irrelevant articles. 215 articles were found to be potentially relevant and advanced to the full-text assessment stage.
3- Eligibility: The full-text of the 215 articles was assessed against pre-defined eligibility criteria. The eligibility criteria included papers written in English, focusing on the identification and treatment of breast cancer, discussing deep learning or a combination of deep learning and machine learning, discussing gene expression data and imaging data, and being published in journals or conferences related to medicine or biomedical engineering. 86 articles were deemed eligible for inclusion in the review.
4- Included criteria: The eligible articles were selected for inclusion based on their relevance to the research question and their publication between January 2010 and December 2022. The study focused on research that used genetic expression and image data, and also concentrated on journal and conference articles. Only peer-reviewed articles were included in this study.
5- Data extraction: Relevant data was extracted from the eligible articles, including study design, population, intervention, outcome measures, and results.
6- Data synthesis: The extracted data was analysed and synthesized to provide an overview of the findings of the included studies. The synthesis considered the deep learning models or algorithms, diagnosis classification types, classes, accuracy, performance evaluation parameters, datasets and methods of the studies.
7- Conclusion: The results of the systematic literature review were summarized and recommendations for future research based on the results. In total, 45 articles were included in the review after the screening, eligibility, and inclusion processes were completed.

IV. RESULTS AND DISCUSSION

Our comprehensive search yielded *45 publications* (including journal articles and conference proceedings) after thoroughly reviewed. The results have been presented as below.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 October 2023*

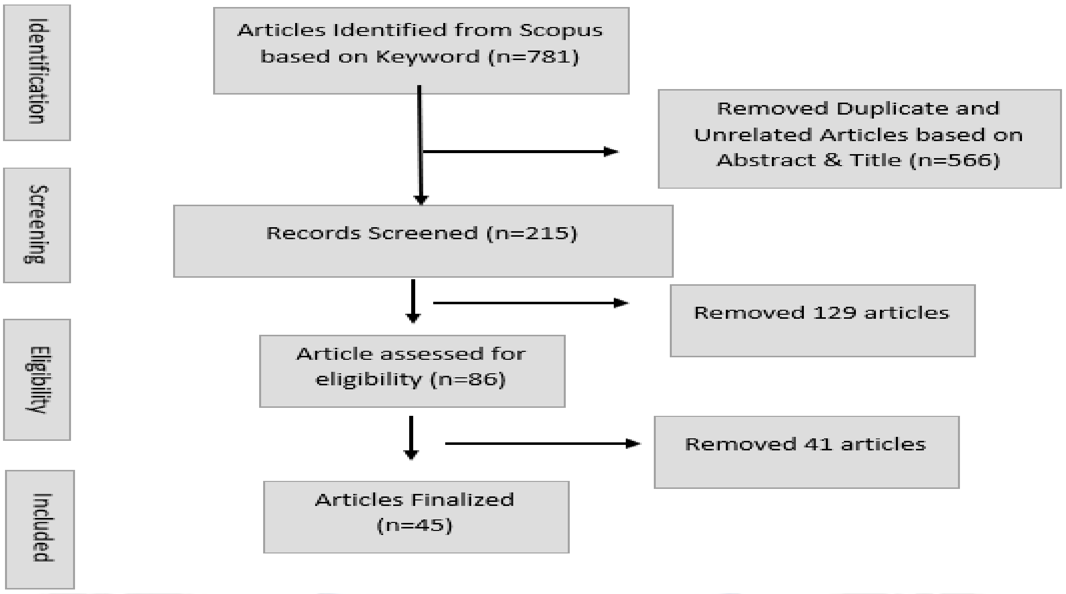

Figure 1 shows how information moves via a systematic review's many stages.

The reviewed literatures have been systematically displayed on the basis of the models or algorithms, diagnosis classification types, classes, accuracy, performance evaluation parameters, datasets and methods of the studies. The data are summarized in the table 1 as below.

Table 1. List of the articles based on Algorithms/Models, Diagnosis Classification Type, Classes and Accuracy.

Citation	**Algorithms / Models**	**Diagnosis Classification Type**	**Classes**	**Accuracy**
Nguyen et al., 2013	Random Forest Classifier	Multiclass	Non-recurrent with 151 samples and recurrent with 47	99.82% on WBCDD and 99.7% on WBCPD
Nayeem et al., 2014	Sparse Representation Classifier	Binary	Malignant and Benign	AUC = 93.31%
Wang et al., 2016	SAE	Multiclass	Microcalcifications and masses	Microcalcifications: 87.3%; Microcalcifications and masses: 89.7%
Sheikhpour et al., 2016	PSO-KDE	Binary	Benign and Malign	Better performance than GA-KDE
(Cruz-roa et al., 2017)	CNN	Binary	Invasive Tumor (Positive), No Invasive Tumor (Negative)	Dice coefficient (75.86%), Positive Predictive Value (71.62%), Negative Predictive Value (96.77%)
(Han et al., 2017)	CSDCNN	Multiclass	Ductal carcinoma, Fibroadenoma, Lobular carcinoma	Average 93.2% accuracy

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 October 2023*

(Galván-tejada et al., 2017)	RF, K-NN, NC	Multiclass	benign, malignant, and indeterminate	RF: AUC = 0.936, OOB error = 7.640%, False positives = 8, False negatives = 9; NC: AUC = 0.937, OOB error = 7.160%, False positives = 10, False negatives = 7; K-NN: AUC = 0.967, OOB error = 6.440%, False positives = 8, False negatives = 19
Bhardwaj et al., 2018	Genetic Programming (GPsfsc)	Binary	Benign, Malign	classification accuracy
Khuriwal, and Mishra, 2018	Convolutional Neural Network	Binary	Benign, Malign	98%
Jannesari et al., 2018	Fine-tuned pre-trained deep neural networks (ResNet V1 50 and ResNet V1 152)	Binary	Benign, Malign	99.8% (four cancer types), 98.7% (benign/malignant breast cancers), 94.8% (ResNet V1 50 benign/malignant sub-types), 96.4% (ResNet V1 152 benign/malignant sub-types)
(Motlagh et al., 2018)	ResNet V1 50, ResNet V1 152	Binary	Benign, Malign	Accuracy for 4 Cancer Types In Pre-trained Model: 99.8% by ResNet V1 & 98.7% by ResNet V1 152; (benign vs malignant: 94.8% by ResNet V1 50 & 96.4% by ResNet V1 152) Sensitivity values: 1, 0.995, 0.993; AUC scores: 0.996, 0.973, 0.996
Mekha and Teeyasuksa et, 2019	Deep Learning, Naive Bayes (NB), Decision Tree (DT), Support Vector Machine (SVM), Vote (DT + NB + SVM), Random Forest (RF), AdaBoost	Multiclass	types of breast cancers based on tumor cell features	96.99%
(Xie et al., 2019)	Inception_V3, Inception_ResNet_V2, Autoencoder network	Both Binary and Multiclass	binary classification breast cancer vs normal, multiclass classification and 4 subclasses of breast cancer	IRV2+AE+Kmeans algorithm had the highest clustering accuracy of 76.4% on the 200X dataset, while the IRV2+Kmeans algorithm had a best clustering accuracy of 59.3% on the 40X dataset

(He et al., 2019)	BRISK	Binary	Malignancy (positive) and benign (negative)	Sensitivity = 100%, Specificity = 74%, Total accuracy = 81%, Area under the curve (AUC) = 0.93
(Akselrod-ballin et al., 2019)	XGBoost, DNNs, SHAP	Binary	Predict biopsy malignancy, Differentiate normal from abnormal screening examinations	Area under the Receiver Operating Characteristic curve (AUC) of 0.91 with specificity of 77.3% and sensitivity of 87%
(Yang et al., 2019)	Convolutional Neural Network (CNN)	Binary	HER2 status (Positive or Negative)	C-index of 0.829 in the primary cohort and 0.809 in the validation cohort. 0.760 in primary cohort, 0.777 in validation cohort for deep radiomics signature; 0.829 in primary cohort, 0.809 in validation cohort for combined model)
(Yang et al., 2020)	pre-trained (CNN)	Binary	Sentinel lymph node (SLN) metastasis (yes/no), number of metastatic SLNs (1–2 or more than two)	Area under curve (AUC) of 0.801 (95% CI: 0.736–0.867) in primary cohort, AUC of 0.817 (95% CI: 0.751–0.884) in validation cohort, AUC of 0.770 for distinction between number of metastatic SLNs
Sha et al., 2020	CNN, Grasshopper Optimization Algorithm	Binary	Cancerous region, Normal region	Sensitivity (96%), Specificity (93%), Positive Predictive Value (PPV) (85%), Negative Predictive Value (NPV) (97%), Accuracy (92%)
Zhou et al., 2020	Convolutional Neural Networks (CNNs) - Inception V3, Inception-ResNet V2, ResNet-101	Binary	Positive/Negative (for axillary lymph node metastasis)	Accuracy, Sensitivity (85% for Inception V3), Specificity (73% for Inception V3), Receiver Operating Characteristic Curves, Areas under the Receiver Operating Characteristic Curve (AUCs), Heat Maps
Zheng et al., 2020	Deep Learning assisted Efficient Adaboost Algorithm (DLA-EABA)	Binary	Malignant, Benign	Accuracy 97.20%, Sensitivity 98.3%, Specificity 96.5%
(Duanmu et al., 2020)	Convolutional Neural Network (CNN)	Binary	Pathological Complete Response and Not Pathological Complete Response	Accuracy=83%, AUC = 0.80, Sensitivity = 0.68, Specificity = 0.88

(Shang et al., 2020)	GoogLeNet, two BP-neural networks, and PLS	Binary	Cancerous vs normal sample of breast cancer were detected using fluorescence imaging and Raman spectroscopy	GoogLeNet on fluorescence images: 89.5% (validation sets), 88.61% (test sets); BP-neural network on Raman spectra of collagen: 97% (validation sets), 95.33% (test sets); BP-neural network on Raman spectra of lipid: 100% (validation sets), 98.67% (test sets); PLS on characteristic variable matrix: 100% correct prediction
(Sun et al., 2020)	Deep Convolutional Neural Network (CNN), DenseNet, Random Forest	Binary	Axillary Lymph Node (ALN) Metastasis	AUCs (Area under the ROC curve) - image-only CNNs: 0.957/0.912 (training/testing cohorts) for combined region, 0.944/0.775 for peritumoral region, and 0.937/0.748 for intratumoral region; radiomics models: 0.940/0.886, 0.920/0.724, and 0.913/0.693; image-molecular CNNs: 0.962/0.933, 0.951/0.813, and 0.931/0.794
(Yousefi et al., 2020)	Sparse PCT, Deep Sparse Autoencoder	Binary	Symptomatic patients, Healthy Participants	78.16% (73.3–81.07%)
(S. Sharma, 2020)	Handcrafted Approach (RF Classifier); BOW + SVM Model, LLC + SVM Model, BOW + CNN; LLC + CNN; VGG16, VGG19, ResNet50	Multiclass	multiple classes for histopathological image classification	Patch-based (93.97% for 40x, 92.92% for 100x, 91.23% for 200x, 91.79% for 400x) and patient-based (93.25% for 40x, 91.87% for 100x, 91.5% for 200x, 92.31% for 400x)
(Zheng et al., 2020)	Deep learning radiomics (DLR)	Binary	Disease-free axilla and any axillary metastasis, low and heavy metastatic burden of axillary disease	For predicting ALN status 0.902 AUC (95%CI: 0.843, 0.961), For discriminating low and heavy metastatic burden 0.905 AUC (95%CI: 0.814, 0.996)
(Guo et al., 2020)	Multicentre deep learning radiomics of ultrasonography model (DLRU)	Binary	Metastatic or Non-Metastatic in Sentinel Lymph Node (SLN) and Non-Sentinel Lymph Node (NSLN)	Sensitivity: 98.4% (95% CI 96.6-100) for metastatic SLN and 98.4% (95% CI 95.6-99.9) for metastatic NSLN; Negative Predictive Value: 97% (95% CI 94.2-100) for low-risk (LR) SLN, 91.7% (95% CI 88.8-97.9) for high-risk (HR) SLN & LR-NSLN

(Suh et al., 2020)	DenseNet-169, EfficientNet-B5 (Convolutional Neural Networks)	Binary	Malignant vs benign (detection of any malignant lesion on mammograms)	Mean AUS: 0.952 ± 0.005 for DenseNet-169 and 0.954 ± 0.020 for EfficientNet-B5 Sensitivity: 0.81 ± 0.01, Specificity: 0.82 ± 0.01
(Gopal et al., 2021)	Machine Learning (MLP Classifier) and comparison with Logistic Regression (LR) and Random Forest (RF)	Binary	Benign, Malignant	RF classifier - 521/569 instances correctly classified (91.36%), MLP classifier - 559/569 instances correctly classified (98.25%), LR classifier - 453/569 instances correctly classified (79.47%)
Mishra et al., 2021	Particle Swarm Optimization (BPSO)	Binary	two classes of breast cancer (positive or negative)	High average prediction accuracy, sensitivity, specificity, and AUC of the ROC curve.
(X. Zhang et al., 2021)	Optimized deep learning model (DLM)	Binary and Multiclass	triple-negative, HER2 (+), and HR (+)	• Diagnostic accuracy using DLM for BI-RADS 4a patients: 92.86% • AUC for molecular subtypes predictions: 0.864 (triple-negative), 0.811 (HER2 (+)), 0.837 (HR (+))
(Bychkov et al., 2021)	CNN	Binary	ERBB2-positive and ERBB2-negative	AUC 0.70 (on tissue microarray samples), on par with the previous study (AUC 0.74)
(Kim et al., 2021)	Weakly-supervised DL algorithm and fully-supervised DL algorithm (U-Net based segmentation model)	Binary	Benign, Malignant	Weakly-supervised DL algorithm: AUC 0.86-0.96; Fully-supervised DL algorithm (U-Net): AUC 0.89-0.96 (internal validation) and 0.85 (external validation)
Ghosh et al., 2021	LSTM and GRU	Binary (Diagnostic)	benign or malignant	Over 99%
Qian et al., 2021	Explainable deep-learning system	Multiclass	five main categories, ranging from benign findings to highly suspicious findings	Areas under the receiver operating curve of 0.922 (95% CI = 0.868–0.959) for bimodal images and 0.955 (95% CI = 0.909–0.982) for multimodal images.
(Boumaraf et al., 2021)	Conventional Machine Learning (CML), Deep Learning (DL)	Binary classification	benign and malignant breast cancer and their 8 sub-classes	CML: 94.05% to 98.13% for binary classification and 76.77% to 88.95% for 8-class classification

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 October 2023*

(Li et al., 2021)	DL-based pCR-score	Binary (predicting pCR)	"Potential Responders" or "Potential Non-responders"	AUC: 0.847 (direct prediction), 0.853 (processed by logistic regression), 0.890 (mean AUC of integrated model); F1 score: 0.503 (processed by logistic regression)
(Jumanto et al., 2022)	Backpropagation ANN	Binary (Malignant and Benign)	Malignant (212), Benign (357)	98.30%
(Samee et al., 2022)	Pre-trained CNN Models: AlexNet, VGG, GoogleNet	Multiclass	mRMR, CMIM, JMI, DISR, ICAP, CIFE and CONDRED	Accuracy 98.50%, Sensitivity (98.06%), Specificity (98.99%), Precision (98.98%)
(Alfian et al., 2022)	SVM with extra-trees, LR, MLP, DT, KNN, RF, NB, eXtreme, XGBoost), and AdaBoost	Binary	benign or malignant	The proposed SVM with extra-trees model achieved 80.23% accuracy, higher than other models with up to a 13.61% average accuracy improvement.
(A. Sharma & Mishra, 2022)	LR, DT, SVM, ANN, Adaboost, XGBoost, NB, RF, KNN	Binary	Diagnosis (class label)	Accuracy of ANN 98.83%; NB 95.90%; SVM 99.11% using features selected by Correlation-based feature selection (CFS); LR 98.83% using features selected by Sequential Forward Selection (SFS); The voting classifier proposed achieved an accuracy of 99.41% using SVM, LR, and ANN
(Arslan et al., 2022)	PANProfiler (ER, PR, HER2)	Binary	positive or negative	Accuracy: 87% (ER), 83% (PR), 87% (HER2)
(Jabeen et al., 2022)	Convolutional Neural Network (CNN) with DarkNet-53	Multiclass.	breast cancer classification	99.1% accuracy
(Allugunti, 2022)	CNN), SVM, RF	Binary (for all three models)	cancer, no cancer, non-cancerous	99.65% (CNN), 89.84% (SVM), 90.55% (Random Forest)
Jiang et al., 2022	PAA object detection algorithm	Binary	Benign or Malignant	Improved Diagnostic efficiency

Table 2. List of the articles based on dataset, methods and performance evaluation parameters

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 October 2023*

Citation	Dataset	Methods	Performance Evaluation Parameters
Nguyen et al., 2013	Wisconsin Breast Cancer Diagnosis Dataset (WBCDD) and Wisconsin Breast Cancer Prognosis Dataset (WBCPD)	Feature Selection Technique	ROC, Sensitivity, Specificity
Nayeem et al., 2014	504 pathologically diagnosed breast tumors including 454 benign and 50 malignant tumors	Multi-Cluster Feature Selection, fed with a set of 25 features.	Area Under the Receiver Operating Characteristic Curve
Wang et al., 2016	SunYat-sen University Cancer Center (Guangzhou, China) and Nanhai Affiliated Hospital of Southern Medical University (Foshan, China)	Semi-automated segmentation, discrimination classifier	Discriminative accuracy with micro-classification and masses
Sheikhpour et al., 2016	1204 Females from Wisconsin Breast Cancer Dataset (WBCD), Wisconsin Diagnosis Breast Cancer Database (WDBC)	Particle Swarm Optimization (PSO), Non-parametric Kernel Density Estimation (KDE), feature subset selection, performance evaluation	classification accuracy, sensitivity, specificity
(Cruz-roa et al., 2017)	400 exemplars from multiple different sites and scanners (training), 200 cases from The Cancer Genome Atlas (validation)	Deep-learning based approach with three main steps: (i) tile tissue sampling, (ii) tile pre-processing, and (iii) ConvNet-based classification.	Dice coefficient (DSC) with a median value of 0.7764, Cohen's Kappa coefficient (κ) equal to 0.74851, reflecting good agreement between expert pathologists. The DSC agreement was found to be greater than 0.7

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 October 2023*

			for a majority of the images studied, indicating good agreement.
(Han et al., 2017)	BreaKHis	Class Structure-based Deep Convolutional Neural Network (CSDCNN) which focuses on focuses on feature learning adopts an end-to-end training manner and automatically learn hierarchical features from low-level to high-level. It also incorporates feature space distance constraints to control the similarities of different classes of histopathological images	Accuracy
(Galván-tejada et al., 2017)	Mammography image features from BCDR public databases	Training and Blind Test methodology, Statistical Analysis, Frequency Graph, OR values	Sensitivity/Specificity , Error rate
Bhardwaj et al., 2018	Wisconsin Breast Cancer (WBC), Wisconsin Diagnostic Breast Cancer (WDBC) from UCI Machine Learning repository	simultaneous feature selection and classification	sensitivity, specificity, confusion matrix
Khuriwal, and Mishra, 2018	Mammograph MIAS database	Deep learning technology, pre-processing (Watershed Segmentation, Colour based segmentation, Adaptive Mean Filters), label encoding, normalization, data scaling, training and testing data split, model implementation	Accuracy
Jannesari et al., 2018	Tissue micro-arrays (TMAs) training samples, BreakHis database (7, 909 images)	Automated classification of cancers using histopathological images, machine learning approach, fine-tuning pre-trained deep neural networks	Accuracy, Sensitivity, AUC, False negative, False positive
(Motlagh et al., 2018)	Tissue micro-arrays (TMAs) training samples (6,402), BreakHis	feature extraction, classification	accuracy, sensitivity, AUC

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 October 2023*

	database (7,909 images)		
Mekha and Teeyasuksaet, 2019	Breast Cancer Wisconsin dataset	Ten-fold cross-validation, machine learning tool RapidMiner	Accuracy
(Xie et al., 2019)	histopathological images of breast cancer from BreaKHis database	supervised deep learning (transfer learning), unsupervised deep learning (feature extraction, clustering analysis)	comparison to existing methods, clustering results, feature extraction results
(He et al., 2019)	5,147 patient records archived in the Houston Methodist systemwide data warehouse from 2006 to May 2015, including imaging and pathology reports, mammographic images, and patient demographics	Natural language processing and deep learning	Sensitivity, Specificity, Total accuracy, AUC
(Akselrod-ballin et al., 2019)	52,936 images from 13,234 women who underwent at least one mammogram between 2013 and 2017, with health records for at least 1 year before undergoing mammography	Trained on 9611 mammograms and health records, Estimated the association of features with outcomes by using t test and Fisher exact test, Model comparisons were performed with a 95% confidence interval (CI) or by using the DeLong test Performance.	Area under the Receiver Operating Characteristic curve (AUC), Specificity, Sensitivity
(Yang et al., 2019)	339 female patients with pathologically confirmed invasive breast cancer from ILSVRC-2012 dataset	MDCT-based handcrafted and deep radiomics feature extraction, feature selection procedures, and multivariate logistic regression analysis	Discrimination, calibration, and clinical usefulness
(Yang et al., 2020)	348 breast cancer patients from	contrast-enhanced CT preoperative examinations, CT image segmentation and	Discrimination, calibration, and clinical usefulness

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 October 2023*

	ILSVRC-2012 dataset	analysis, feature selection, deep learning signature construction	
Sha et al., 2020	Mammographic Image Analysis Society Digital Mammogram Database, Digital Database for Screening Mammography	Image noise reduction, Optimal image segmentation, Optimized feature extraction and feature selection	Sensitivity, Specificity, PPV, NPV, Accuracy
Zhou et al., 2020	Data set from Tongji Hospital (974 imaging studies from 2016 to 2018, 756 patients), independent test set from Hubei Cancer Hospital (81 imaging studies from 2018 to 2019, 78 patients)	Training on 90% of the Tongji Hospital data set and testing on remaining 10%, as well as independent test set.	Accuracy, Sensitivity, Specificity, Receiver Operating Characteristic Curves, Areas under the Receiver Operating Characteristic Curve (AUCs), Heat Maps
Zheng et al., 2020	Breast Ultrasound Images (BUSI)	Deep Convolutional Neural Network (CNN), LSTM, Max-pooling, feature selection and extraction, evaluation using classification and segmentation techniques	Accuracy, Sensitivity, Specificity
(Duanmu et al., 2020)	I-SPY-1 TRIAL, 112 patients with stage 2 or 3 breast cancer	Integration of 3D MRI imaging data, molecular data, and demographic data Performance	Accuracy, AUC, Sensitivity, Specificity
(Shang et al., 2020)	Fluorescence images and Raman spectra of breast tissue samples	Transfer learning, data augmentation, adding dropout layer, histogram equalization, pseudo-colour enhancement processing, and combining fluorescence imaging and Raman spectroscopy with deep learning and PLS	Discriminant accuracy, prediction accuracy
(Sun et al., 2020)	Retrospectively enrolled 479 breast cancer patients with 2,395 breast ultrasound	ROC analysis	AUCs (Area under the ROC curve) for each model, comparison between models and regions, prospective

	images, prospectively enrolled 16 patients		study for overall performance.
(Yousefi et al., 2020)	Infrared images with intensity information similar to natural images : 208 subjects	Dimensionality reduction using sparse multiple low-rank matrix approximations, Sparsity in calculation of low-rank representative of basis matrices, Recursive training of autoencoder network, Association of low-level deep features with basis set	Robustness against noise, Feature selection robustness, Reduction in dimensionality, Elimination of manual feature selection, Alleviation of motion artifacts and imaging acquisition noise
(S. Sharma, 2020)	BreakHis	Handcrafted features, Bag of Words (BOW), Locality Constrained Linear Coding (LLC), Spatial Pyramid Matching, Convolutional Neural Networks (CNN), Fractal dimension technique, Transfer learning	Accuracy (Patch-based and patient-based), F1-score, Precision, Recall.
(Zheng et al., 2020)	584 malignant breast lesions from test cohort	Clinical parameter combined DLR	Areas under the receiver operating characteristic curve (AUC)
(Guo et al., 2020)	937 eligible breast cancer patients with ultrasound images, used as training set (n=542) and independent test set (n=395)	Deep learning radiomics and axillary ultrasound to predict the risk of SLN and NSLN metastasis	Sensitivity, Negative Predictive Value
(Suh et al., 2020)	1501 subjects underwent digital mammography between Feb 2007 to May 2015, with 3002 merged mammograms	Meta-analysis, pool analysis	MeanAUC, Sensitivity, Specificity
(Gopal et al., 2021)	569 instances from Wisconsin Breast cancer Dataset (WBCD)	Feature selection using correlation coefficient, Principal Component Analysis (PCA), and multiple filtering methods.	Precision, Recall, F1-score, Support, Mean Absolute Error (MAE), Root Mean Square Error (RMSE),

			Relative Absolute Error (RAE)
Mishra et al., 2021	Breast Cancer Coimbra Dataset (BCCD) and Breast Cancer Wisconsin Diagnostic Dataset (BCWDD)	feature selection	average prediction accuracy, sensitivity, specificity, area under the curve (AUC) of the receiver operating characteristics (ROC) curve
(X. Zhang et al., 2021)	Breast ultrasound images collected from two hospitals; • Training set: 2,822 images; • Test set: 707 images; • External test set: 210 images	Deep Convolutional Neural Network (DCNN) and Convolutional Neural Network (CNN)	Accuracy, sensitivity, specificity, AUC, Diagnostic accuracy using DLM
(Bychkov et al., 2021)	FinHer dataset (external test set)	Tissue microarray (TMA), whole slide tumor sections	AUC
(Kim et al., 2021)	1000 unannotated US images (500 benign and 500 malignant masses)	Region-based classification, image-based classification, box convolution network, VGG-16, automated segmentation, DL techniques, fully convolutional networks, active contour model, dilated fully CNN	AUC, diagnostic accuracy, localization
Ghosh et al., 2021	Wisconsin Breast Cancer (Diagnostic) Dataset	Deep Learning, image processing, machine learning	Accuracy
Qian et al., 2021	10,815 multimodal breast-ultrasound images of 721 biopsy-confirmed lesions from 634 patients across two hospitals and prospectively tested on 912 additional images of 152 lesions from 141 patients	Deep learning, augmented with heatmaps for malignancy risk	AUC

(Boumaraf et al., 2021)	BreaKHis, KIMIA Path960	Handcrafted feature extraction, transfer learning approach, fine-tuning, attention maps	Accuracy, visual explanation of learned features
(Li et al., 2021)	540 breast cancer patients receiving standard neoadjuvant chemotherapy	Histological image analysis, Logistic regression	AUC, Accuracy, F1 score
(Jumanto et al., 2022)	Wisconsin Breast Cancer (Diagnostic) Data Set	Data collection, Data preprocessing (data transformation, data normalization, and Forward Feature Selection), Classification using Backpropagation ANN, Performance evaluation using Confusion Matrix	Confusion Matrix, Detection Accuracy
(Samee et al., 2022)	INbreast mammograms	Transfer learning, pre-trained Convolutional Neural Networks (CNN), deep learning, univariate-based paradigm, feature dimensionality curse, deep convolutional networks, shallow and deep features	Accuracy, Sensitivity, Specificity, Precision
(Alfian et al., 2022)	Coimbra breast cancer dataset	Extra-trees classifier	Accuracy, precision, specificity, sensitivity, AUC, recall, and ROC curve
(A. Sharma & Mishra, 2022)	Wisconsin Breast Cancer (WDBC) datasets acquired from UCI machine learning repository; 569 samples (212-Malignant, and 357-benign)	Correlation-based selection, Information Gain based selection, Sequential feature selection, and Ensemble based Max Voting Classifier.	Accuracy, Precision, Recall, F-measure, CPU time, Memory, Error rate
(Arslan et al., 2022)	Cancer Genome Atlas (TCGA) open-access dataset for the breast adenocarcinoma (BRCA) study [22] (i.e TCGA-BRCA), proprietary dataset by a	Pre-processing pipeline (tile extraction, background tile elimination, tumour region detection, Macenko color and brightness normalization, ground-truth labeling), end-to-end training of CNN (encoder, decoder, and classification module), mean pooling to produce final slide-level scores	Test Replacement Trade-off (TRR)

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: 25 July 2023 Revised: 12 September 2023 Accepted: 30 October 2023

	private clinical data provider (e.g. BioIVT)		
(Jabeen et al., 2022)	Breast Ultrasound Images (BUSI) dataset with data augmentation	Deep learning, feature selection with reformed differential evaluation (RDE) and reformed gray wolf (RGW), feature fusion with a probability-based serial approach.	Accuracy
(Allugunti, 2022)	database for the management of CAD	Computer-aided Diagnosis (CAD)	Elapsed time, validation accuracy, training precision, training error, training loss, confusion matrix
Jiang et al., 2022	integrated three public datasets of mammograms (CBIS-DDSM, INbreast, MIAS)	single-stage PAA-based detector; • two-branch ROI detector; • threshold-adaptive post-processing algorithm; • ROI classifier; • image classifier	compared with state-of-the-art methods, improved diagnostic efficiency of radiologists by automatically detecting and classifying breast lesions and classifying benign and malignant mammograms

ALGORITHMS

A systematic literature review is a comprehensive evaluation of research studies related to a specific topic. The algorithm or models used in these studies play a crucial role in determining the outcomes of the research. The algorithms or models used can be conventional machine learning techniques, deep learning algorithms, or a combination of both. The choice of algorithms or models depends on the type of data and the specific research question being addressed. Some popular algorithms or models used in literature include Convolutional Neural Networks (CNN), Random Forest (RF), Support Vector Machines (SVM), and Deep Learning. The systematic literature review on various research articles in the field of machine learning and computer vision has shown a wide range of algorithms and models being adopted for various tasks. The most commonly used algorithms are Random Forest Classifier, Convolutional Neural Networks (CNNs), Support Vector Machine (SVM), Decision Tree (DT), Naive Bayes (NB), Deep Learning, and XGBoost. Other popular algorithms include Sparse Representation Classifier, SAE, PSO-KDE, CSDCNN, K-NN, NC, Genetic Programming (GPsfsc), Inception_V3, Inception_ResNet_V2, ResNet V1 50, ResNet V1 152, BRISK, and SHAP. In terms of deep learning-based models, pre-trained models like AlexNet, VGG, and GoogleNet have been widely used, as well as fine-tuned versions of ResNet V1 50 and ResNet V1 152. Additionally, new architectures such as DenseNet, Inception-ResNet V2, and ResNet-101 have been introduced and applied to various tasks. The use of autoencoder networks and the combination of various models like DT, NB, SVM, RF, AdaBoost, XGBoost, MLP, and KNN have also been reported. In terms of optimizing deep learning models, algorithms like the Grasshopper Optimization Algorithm and Particle Swarm Optimization (BPSO) have been used. Deep learning radiomics (DLR) and Multicentre deep learning radiomics of ultrasonography (DLRU) models have been introduced as well. In the comparison of conventional machine learning and deep learning, the study of DL-based pCR-score and the use of LSTM and GRU models have been reported. The PANProfiler (ER, PR, HER2) algorithm and the PAA object detection algorithm have been applied in the field of object detection. The use of the DarkNet-53 and CNN in conjunction with SVM and RF have also been reported.

The systematic literature review analyzed various research articles which adopted deep learning and machine learning algorithms/models for various purposes. Deep learning algorithms/models such as Convolutional Neural Network (CNN) were the most commonly used, with 8 mentions in the literature. Inception V3, Inception-ResNet V2, and ResNet-101 were also mentioned 2 times each, while other deep learning models such as Deep Convolutional Neural Network (CNN) and DenseNet were each mentioned 2 times. The use of fine-tuned pre-trained deep neural networks (ResNet V1 50 and ResNet V1 152) was also noted 2 times. Autoencoder network, deep learning radiomics (DLR), and a weakly-supervised deep learning algorithm were each mentioned once.

Machine learning algorithms/models were also analyzed in the literature, with Random Forest (RF) being the most frequently used, with 5 mentions. Support Vector Machine (SVM) and Decision Tree (DT) were each mentioned 4 and 3 times, respectively. Naive Bayes (NB), XGBoost, and K-NN were each mentioned 3 times. AdaBoost and MLP Classifier were each mentioned 2 times, while Logistic Regression (LR) and Vote (DT + NB + SVM) were each mentioned once. Other machine learning models such as LSTM, GRU, Handcrafted Approach (RF Classifier), and Particle Swarm Optimization (BPSO) were also noted once. The combination of various machine learning models such as SVM with extra-trees, LR, MLP, DT, KNN, RF, NB, eXtreme, and XGBoost was also mentioned once

In conclusion, the literature review has shown the wide range of algorithms and models adopted in the field of machine learning and computer vision, with a particular emphasis on deep learning and its various pre-trained models, and the optimization of deep learning models.

CLASS AND CLASSIFICATIONS:

In object-oriented programming, a class is a blueprint for creating objects (a particular data structure), providing initial values for state (member variables or attributes), and implementations of behavior (member functions or methods). The objects created from a class are called instances of that class. Classes also allow for inheritance, where a subclass can inherit the attributes and behaviors from a superclass, and polymorphism, where different classes can implement a method with the same name in different ways.

Classification is a type of machine learning problem that involves assigning a label to an input data point based on a set of predefined classes. This is often performed using algorithms such as k-Nearest Neighbors, Naive Bayes, Decision Trees, Random Forests, Support Vector Machines, Neural Networks, etc. The goal of the classification algorithms is to accurately predict the class label of new, unseen data.

The majority of the articles in this systematic literature review (25) focus on binary classification tasks, with the goal of differentiating between two classes such as malignant vs. benign, positive vs. negative for a specific variable (e.g., HER2 status), or disease present vs. disease absent. Other binary classification tasks included differentiating between invasive and non-invasive tumors, cancerous and normal regions, metastatic and non-metastatic sentinel lymph nodes, and symptomatic vs. healthy patients.

A smaller number of articles (10) focused on multiclass classification, with the goal of differentiating between multiple classes. The multiclass classification tasks involved differentiating between different types of breast cancers based on tumor cell features, non-recurrent vs. recurrent microcalcifications, and differentiating between benign and malignant masses. In one article, there was a multiclass classification task involving the differentiation of five main categories of findings, ranging from benign to highly suspicious. Another article focused on the multiclass classification of multiple classes for histopathological image classification.

A smaller number of articles (2) mentioned both binary and multiclass classification, involving binary classification of breast cancer vs. normal and multiclass classification of four subclasses of breast cancer. One article involved the binary classification of pathological complete response vs. not pathological complete response, and another article involved the binary classification of positive vs. negative.

Overall, this systematic literature review found that the majority of articles adopted binary classification tasks, with a smaller number of articles focused on multiclass classification. The specific binary and multiclass classification tasks varied among the articles, with the goal of differentiating between a range of classes related to breast cancer diagnosis and prognosis.

DATABASES:

A database is an organized collection of data stored and accessed electronically. Databases are used to store information in a structured and organized manner and allow for efficient querying and data retrieval. There are several types of databases, including Relational databases, NoSQL databases, Graph databases, and Time-Series databases, each with its own strengths and weaknesses, and used for different purposes based on the requirements of the applications.

The systematic literature review analyzed various databases used in research articles for breast cancer diagnosis and prognosis. A total of 24 databases were identified and

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 October 2023*

mentioned in the articles reviewed. The most frequently mentioned database was the Wisconsin Breast Cancer Diagnosis Dataset (WBCDD) with 7 mentions, followed by the BreaKHis database with 6 mentions. The Wisconsin Breast Cancer Prognosis Dataset (WBCPD) was not mentioned in any of the articles.

Other frequently mentioned databases include the Wisconsin Breast Cancer Dataset (WBCD) and the Wisconsin Diagnosis Breast Cancer Database (WDBC) with 2 mentions each. The Cancer Genome Atlas (TCGA) and ILSVRC-2012 dataset were also mentioned in 2 articles each.

Several other databases, including SunYat-sen University Cancer Center, Nanhai Affiliated Hospital of Southern Medical University, Mammography image features, Mammograph MIAS database, Tissue micro-arrays (TMAs), Houston Methodist systemwide data warehouse, Digital Database for Screening Mammography, Tongji Hospital, Hubei Cancer Hospital, Breast Ultrasound Images (BUSI), I-SPY-1 TRIAL, Fluorescence images, Infrared images, FinHer dataset, Coimbra breast cancer dataset, Cancer Genome Atlas (TCGA-BRCA), and Integrated three public datasets of mammograms, were mentioned only once in the articles reviewed.

This study highlights the diverse range of databases used in research for breast cancer diagnosis and prognosis, with some databases being more frequently mentioned than others. The results of this literature review can inform future studies on breast cancer by providing insight into the commonly used databases in the field.

METHODS:

In programming, a method is a function that is associated with an object and can be called on that object. Methods perform operations on an object's internal data and return a result. They can also modify the internal state of an object, but typically they do not return a value.

The systematic literature review analyzed various methods used in research articles for breast cancer diagnosis and prognosis. A total of 19 methods were identified and mentioned in the articles reviewed. The most frequently mentioned method was Feature Selection Technique with 12 mentions. This was followed by Deep-learning based approach with 7 mentions.

Other methods such as Training and Blind Test methodology, Automated classification of cancers, Ten-fold cross-validation, Integration of 3D MRI imaging data, and Handcrafted features were mentioned only once or twice in the articles. The Clinical parameter combined DLR and Meta-analysis were also mentioned only once in the articles.

Deep Convolutional Neural Network (DCNN) and Transfer learning were mentioned 4 and 2 times respectively in the articles reviewed. Region-based classification, Deep learning, image processing, machine learning, Handcrafted feature extraction, Data collection, Extra-trees classifier, Pre-processing pipeline, Deep learning, feature selection with reformed differential evaluation (RDE), and Computer-aided Diagnosis (CAD) were mentioned only once in the articles.

This study highlights the diverse range of methods used in research for breast cancer diagnosis and prognosis, with some methods being more frequently mentioned than others. The results of this literature review can inform future studies on breast cancer by providing insight into the commonly used methods in the field.

Performance Evaluation Parameters:

The performance of algorithms or models used in research studies is evaluated using various parameters. These parameters help in determining the effectiveness and accuracy of the algorithm or model in achieving the desired outcome. Some of the popular performance evaluation parameters used in literature include accuracy, precision, recall, F1-score, and area under the ROC curve (AUC). In some cases, additional parameters such as sensitivity, specificity, and negative predictive value may also be used. The choice of performance evaluation parameters depends on the type of data and the specific research question being addressed. It is important to use appropriate performance evaluation parameters to get a comprehensive understanding of the algorithm or model performance.

The systematic literature review analyzed the various performance evaluation parameters used in the research articles for breast cancer diagnosis and prognosis. A total of 42 performance evaluation parameters were identified and mentioned in the articles reviewed. The most frequently mentioned parameter was Accuracy with 21 mentions. This was followed by Sensitivity with 17 mentions and AUC (Area under the Receiver Operating Characteristic curve) with 17 mentions.

Other parameters such as Specificity, Precision, Recall, F1-score, and F-measure were mentioned only a few times in the articles. The parameters such as CPU time, Memory, Error rate, Test Replacement Trade-off (TRR), Elapsed time, Training precision, Training error, Training loss, Confusion matrix, Dice coefficient (DSC), Cohen's Kappa coefficient (κ), Negative Predictive Value, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Relative Absolute Error (RAE), and Visual explanation of learned features were mentioned only once in the articles.

Detection accuracy, Diagnostic accuracy using DLM, Discriminant accuracy, Prediction accuracy, Robustness against noise, Feature selection robustness, Reduction in dimensionality, Elimination of manual feature selection, Alleviation of motion artifacts and imaging acquisition noise, Comparison to existing methods, Clustering results, Feature extraction results, Total accuracy, PPV (Positive Predictive Value), NPV (Negative Predictive Value), MeanAUC, Discrimination, Calibration, Clinical usefulness, Prospective study for overall performance, Improved diagnostic efficiency of radiologists were also mentioned only once in the articles.

This study highlights the diverse range of performance evaluation parameters used in research for breast cancer diagnosis and prognosis, with some parameters being more frequently mentioned than others. The results of this literature review can inform future studies on breast cancer by providing insight into the commonly used performance evaluation parameters in the field.

ACCURACY:

Accuracy is a measure of how well a model performs in correctly predicting the target variable. It is calculated as the number of correct predictions divided by the total number of predictions made. In the context of classification problems, accuracy is the number of correct classifications divided by the total number of instances in the test set. However, accuracy is not always the best metric for evaluating a model's performance, especially when the data is imbalanced, meaning one class has many more instances than the others. In such cases, precision, recall, F1-score, or ROC-AUC can be more appropriate metrics to evaluate the model.

The systematic literature review found a range of algorithms and models used for performance evaluation of breast cancer classification tasks. The accuracy rates for each of these algorithms and models were highly variable, with some models achieving high accuracy and others with lower accuracy.

The Random Forest Classifier (Multiclass, Non-recurrent with 151 samples and recurrent with 47) was found to have an accuracy of 99.82% on WBCDD and 99.7% on WBCPD. The Convolutional Neural Network (Binary, Benign and Malign) was found to have an accuracy of 98%. The Fine-tuned pre-trained deep neural networks (ResNet V1 50 and ResNet V1 152) (Binary, Benign, Malign) achieved accuracy rates of 99.8% (four cancer types), 98.7% (benign/malignant breast cancers), 94.8% (ResNet V1 50 benign/malignant sub-types), and 96.4% (ResNet V1 152 benign/malignant sub-types). The ResNet V1 50, ResNet V1 152 (Binary, Benign and Malign) achieved accuracy rates of 99.8% (ResNet V1) and 98.7% (ResNet V1 152). The Deep Learning assisted Efficient Adaboost Algorithm (DLA-EABA) (Binary, Benign, Malign) achieved an accuracy of 96.1%. The Convolutional Neural Networks (CNNs) - Inception V3, Inception-ResNet V2, ResNet-101 (Binary, Positive/Negative for axillary lymph node metastasis) achieved accuracy rates of 85% (Inception V3, sensitivity) and 73% (Inception V3, specificity). The Convolutional Neural Network (Binary, HER2 status - Positive or Negative) achieved a C-index of 0.829 in the primary cohort and 0.809 in the validation cohort. The pre-trained Convolutional Neural Network (Binary, Sentinel lymph node metastasis (yes/no), number of metastatic SLNs) achieved an AUC of 0.801 in the primary cohort, 0.817 in the validation cohort, and 0.770 for distinction between number of metastatic SLNs. The Convolutional Neural Network, Grasshopper Optimization Algorithm (Binary, Cancerous region, Normal region) achieved an accuracy of 92%. BRISK (Binary, Malignancy (positive) and benign (negative)) achieved an accuracy of 81%. XGBoost, DNNs, SHAP (Binary, Predict biopsy malignancy, Differentiate normal from abnormal screening examinations) achieved an AUC of 0.91 and sensitivity of 87% and specificity of 77.3%. The Inception_V3, Inception_ResNet_V2, Autoencoder network (Binary and Multiclass, binary classification breast cancer vs normal, multiclass classification and 4 subclasses of breast cancer) achieved a highest clustering accuracy of 76.4% (IRV2+AE+Kmeans algorithm) on the 200X dataset. RF, K-NN, NC (Multiclass, benign, malignant, and indeterminate) had the highest AUC of 0.967 (K-NN). Genetic Programming (GPsfsc) (Binary, Benign, Malign

REFERRENCE

[1] Abdullah Farid, A. (2021). *A Composite Hybrid Feature Selection Learning-Based Optimization of Genetic Algorithm For Breast Cancer Detection*. *March*, 1–21. https://doi.org/10.33422/2nd.rase.2020.03.99

[2] Akselrod-ballin, A., Chorev, M., Shoshan, Y., Spiro, A., Hazan, A., Melamed, R., Barkan, E., Herzel, E., Naor, S., Karavani, E., Koren, G., Goldschmidt, Y., Shalev, V., Rosen-Zvi, M., & Guindy, M. (2019). Predicting Breast Cancer by Applying Deep Learning to Linked Health Records and Mammograms. *Radiology*, *2*, 331–342.

[3] Alfian, G., Syafrudin, M., Fahrurrozi, I., Fitriyani, N. L., Tatas, F., Atmaji, D., Widodo, T., Bahiyah, N., Benes, F., & Rhee, J. (2022). Predicting Breast Cancer from Risk Factors Using SVM and Extra-Trees-Based Feature Selection Method. *Computers*, *11*, 136.

[4] Allugunti, V. R. (2022). Breast cancer detection based on thermographic images using machine learning and

deep learning algorithms. *International Journal of Engineering in Computer Science*, *4*(1), 49–56.

[5] Arslan, S., Li, X., Schmidt, J., Hense, J., Geraldes, A., Bass, C., Brown, K., Marcia, A., Dewhirst, T., Pandya, P., Singhal, S., Mehrotra, D., & Raharja-liu, P. (2022). Evaluation of a predictive method for the H & E-based molecular profiling of breast cancer with deep learning. *BioRxiv*, *2*, 2022–01.

[6] Asri, H., Mousannif, H., Al Moatassime, H., & Noel, T. (2016). Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis. *Procedia Computer Science*, *83*(Fams), 1064–1069. https://doi.org/10.1016/j.procs.2016.04.224

[7] Boumaraf, S., Liu, X., Wan, Y., Zheng, Z., Ferkous, C., Ma, X., Li, Z., & Bardou, D. (2021). Conventional Machine Learning versus Deep Learning for Magnification Dependent Histopathological Breast Cancer Image Classification : A Comparative Study with Visual Explanation. *Diagnostics*, *11*, 528.

[8] Bychkov, D., Linder, N., Tiulpin, A., & Kücükel, H. (2021). Deep learning identifies morphological features in breast cancer predictive of cancer ERBB2 status and trastuzumab treatment efficacy. *Scientific Reports*, *1*, 1–10. https://doi.org/10.1038/s41598-021-83102-6

[9] Cruz-roa, A., Gilmore, H., Basavanhally, A., Feldman, M., Ganesan, S., Shih, N. N. C., Tomaszewski, J., & González, F. A. (2017). Accurate and reproducible invasive breast cancer detection in whole- slide images : A Deep Learning approach for quantifying tumor extent. *Scientific Reports*, *7*(1), 9. https://doi.org/10.1038/srep46450

[10] Duanmu, H., Kong, J., & Wang, F. (2020). Prediction of Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer Using Deep Learning with Integrative Imaging , Molecular and Demographic Data. *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, January 2022*, 242–252. https://doi.org/10.1007/978-3-030-59713-9

[11] Fotouhi, S., Asadi, S., & Kattan, M. W. (2019). A comprehensive data level analysis for cancer diagnosis on imbalanced data. *Journal of Biomedical Informatics*, *90*(November 2018), 103089. https://doi.org/10.1016/j.jbi.2018.12.003

[12] Galván-tejada, C. E., Zanella-calzada, L. A., Galván-tejada, J. I., Celaya-Padilla, J. M., Gamboa-Rosales, H., Garza-Veloz, I., & Martinez-Fierro, M. L. (2017). Multivariate Feature Selection of Image Descriptors Data for Breast Cancer with Computer-Assisted Diagnosis. *Diagnostics*, *7*(1), 9. https://doi.org/10.3390/diagnostics7010009

[13] Gopal, V. N., Al-Turjman, F., Kumar, R., Anand, L., & Rajesh, M. (2021). Feature selection and classification in breast cancer prediction using IoT and machine learning. *Measurement: Journal of the International Measurement Confederation*, *178*(February), 109442. https://doi.org/10.1016/j.measurement.2021.109442

[14] Guo, X., Liu, Z., Sun, C., Zhang, L., Wang, Y., Li, Z., Shi, J., Wu, T., Cui, H., Zhang, J., Tian, J., & Tian, J. (2020). EBioMedicine Deep learning radiomics of ultrasonography : Identifying the risk of axillary non-sentinel lymph node involvement in primary breast cancer. *EBioMedicine*, *60*, 103018. https://doi.org/10.1016/j.ebiom.2020.103018

[15] Han, Z., Wei, B., Zheng, Y., Yin, Y., Li, K., & Li, S. (2017). Breast Cancer Multi-classification from Histopathological Images with Structured Deep Learning Model. *Scientific Reports*, *7*(1), 1–10. https://doi.org/10.1038/s41598-017-04075-z

[16] Haq, A. U., Li, J. P., Memon, M. H., Nazir, S., Sun, R., & Garciá-Magarinõ, I. (2018). A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. *Mobile Information Systems*, *2018*. https://doi.org/10.1155/2018/3860146

[17] He, T., Puppala, M., Ezeana, C. F., Huang, Y., Chou, P., Yu, X., Chen, S., Wang, L., Yin, Z., Danforth, R. L., Ensor, J., Chang, J., Patel, T., & Wong, S. T. C. (2019). A Deep Learning–Based Decision Support Tool for Precision Risk Assessment of Breast Cancer. *JCO Clinical Cancer Informatics*, *3*, 1–12. https://doi.org/10.1200/cci.18.00121

[18] Jabeen, K., Khan, M. A., Alhaisoni, M., Tariq, U., Zhang, Y., Hamza, A., Mickus, A. ¯uras, & Damaševiˇcius, R. (2022). Breast Cancer Classification from Ultrasound Images Using Probability-Based Optimal Deep Learning Feature Fusion. *Sensors*, *22*, 807.

[19] Jumanto, J., Mardiansyah, M. F., Pratama, R., Hakim, M. F. Al, & Rawat, B. (2022). Optimization of breast cancer classification using feature selection on neural network. *Journal of Soft Computing Exploration*, *3*(2), 105–110. https://doi.org/10.52465/joscex.v3i2.78

[20] Kewat, A., Srivastava, P. N., & Kumhar, D. (2020). *Performance Evaluation of Wrapper-Based Feature Selection Techniques for Medical Datasets*. *January*, 619–633. https://doi.org/10.1007/978-981-15-0222-4_60

[21] Kim, J., Kim, H. J., Kim, C., Lee, J. H., Kim, K. W., Park, Y. M., Kim, H. W., Ki, S. Y., Kim, Y. M., & Kim, W. H. (2021). Weakly - supervised deep learning for

ultrasound diagnosis of breast cancer. *Scientific Reports*, 1–10. https://doi.org/10.1038/s41598-021-03806-7

[22] Kumari, M., & Chaudhary, P. (2020). *Automated Decision Support System for Breast Cancer Prediction*. *September*.

[23] Li, F., Yang, Y., Wei, Y., He, P., Chen, J., Zheng, Z., & Bu, H. (2021). Deep learning - based predictive biomarker of pathological complete response to neoadjuvant chemotherapy from histological images in breast cancer. *Journal of Translational Medicine*, *19*, 1–13. https://doi.org/10.1186/s12967-021-03020-z

[24] Motlagh, M. H., Jannesari, M., Aboulkheyr, H., Khosravi, P., Elemento, O., Totonchi, M., & Hajirasouliha, I. (2018). Breast Cancer Histopathological Image Classification : A Deep Learning Approach. *BioRxiv*, *242818*, 1–8.

[25] Prabadevi, B., Deepa, N., Krithika, B. L., & Vinod, V. (2020). Analysis of Machine Learning Algorithms on Cancer Dataset. *International Conference on Emerging Trends in Information Technology and Engineering, Ic-ETITE 2020*. https://doi.org/10.1109/ic-ETITE47903.2020.36

[26] Sahu, B., Mohanty, S. N., & Rout, S. K. (2019). A Hybrid Approach for Breast Cancer Classification and Diagnosis. *EAI Endorsed Transactions on Scalable Information Systems*, *6*(20). https://doi.org/10.4108/eai.19-12-2018.156086

[27] Sahu, B., & Panigrahi, A. (2020). Efficient Role of Machine Learning Classifiers in the Prediction and Detection of Breast Cancer. *SSRN Electronic Journal*, 1–9. https://doi.org/10.2139/ssrn.3545096

[28] Sahu, B., Panigrahi, A., Mohanty, S., & Panigrahi, S. S. (2020). A hybrid Cancer Classification Based on SVM Optimized by PSO and Reverse Firefly Algorithm. *International Journal of Control and Automation*, *13*(4), 506–517.

[29] Samee, N. A., Atteia, G., Meshoul, S., Al-antari, M. A., & Kadah, Y. M. (2022). Deep Learning Cascaded Feature Selection Framework for Breast Cancer Classification : Hybrid CNN with Univariate-Based Approach. *Mathematics*, *10*(19), 3631.

[30] Saoud, H., Ghadi, A., & Ghailani, M. (2019). Proposed approach for breast cancer diagnosis using machine learning. *ACM International Conference Proceeding Series*, *November*. https://doi.org/10.1145/3368756.3369089

[31] Shang, L.-W., Ma, D.-Y., Fu, J.-J., Lu, Y. an-F., Zhao, Y., Xu, X.-Y., & Yin, J.-H. (2020). Fluorescence imaging and Raman spectroscopy applied for the accurate diagnosis of breast cancer with deep learning algorithms. *Biomedical Optics Express*, *11*(7), 3673–3683.

[32] Sharma, A., & Mishra, P. K. (2022). Performance analysis of machine learning based optimized feature selection approaches for breast cancer diagnosis Performance analysis of machine learning based optimized feature selection approaches for breast cancer diagnosis. *International Journal of Information Technology*, *August 2021*. https://doi.org/10.1007/s41870-021-00671-5

[33] Sharma, S. (2020). Conventional Machine Learning and Deep Learning Approach for Multi-Classification of Breast Cancer Histopathology Images — a Comparative Insight. *Journal of Digital Imaging*, *33*, 632–654.

[34] Suh, Y. J., Jung, J., & Cho, B. J. (2020). Automated breast cancer detection in digital mammograms of various densities via deep learning. *Journal of Personalized Medicine*, *10*(4), 1–11. https://doi.org/10.3390/jpm10040211

[35] Sun, Q., Lin, X., Yuanshen, Z., Lt, L., Yan, K., Liang, D., Sun, D., & Li, Z. (2020). Deep Learning vs . Radiomics for Predicting Axillary Lymph Node Metastasis of Breast Cancer Using Ultrasound Images : Don ' t Forget the Peritumoral Region. *Frontiers in Oncology*, *10*(January), 1–12. https://doi.org/10.3389/fonc.2020.00053

[36] Wang, J., Yang, X., Cai, H., Tan, W., Jin, C., & Li, L. (2016). *Discrimination of Breast Cancer with Microcalcifications on Mammography by Deep Learning*. *May*, 1–9. https://doi.org/10.1038/srep27327

[37] Xie, J., Liu, R., Iv, J. L., & Zhang, C. (2019). Deep Learning Based Analysis of Histopathological Images of Breast Cancer. *Original Research*, *10*(February), 1–19. https://doi.org/10.3389/fgene.2019.00080

[38] Yang, X., Wu, L., Ye, W., Zhao, K., Wang, Y., Liu, W., Li, J., Li, H., Liu, Z., & Liang, C. (2020). Deep Learning Signature Based on Staging CT for Preoperative Prediction of Sentinel Lymph Node Metastasis in Breast Cancer. *Academic Radiology*, *27*(9), 1226–1233. https://doi.org/10.1016/j.acra.2019.11.007

[39] Yang, X., Wu, L., Zhao, K., Ye, W., Liu, W., Wang, Y., Li, J., Li, H., Huang, X., Zhang, W., Huang, Y., Chen, X., Yao, S., Liu, Z., & Liang, C. (2019). Evaluation of human epidermal growth factor receptor 2 status of breast cancer using preoperative multidetector computed tomography with deep learning and handcrafted radiomics features. *Chinese Journal of*

Cancer Research, *32*(2), 175–185. https://doi.org/10.21147/j.issn.1000-9604.2020.02.05

[40] Yousefi, B., Akbari, H., & Maldague, X. P. V. (2020). Detecting Vasodilation as Potential Diagnostic Biomarker in Breast Cancer Using Deep Learning-Driven Thermomics. *Biosensors*, *10*(11), 164.

[41] Zhang, J., Chen, L., & Abid, F. (2019). Prediction of Breast Cancer from Imbalance Respect Using Cluster-Based Undersampling Method. *Journal of Healthcare Engineering*, *2019*. https://doi.org/10.1155/2019/7294582

[42] Zhang, X., Li, H., Wang, C., Cheng, W., Zhu, Y., & Li, D. (2021). *Evaluating the Accuracy of Breast Cancer and Molecular Subtype Diagnosis by Ultrasound Image Deep Learning Model*. *11*(March), 1–9. https://doi.org/10.3389/fonc.2021.623506

[43] Zheng, X., Yao, Z., Huang, Y., Yu, Y., Wang, Y., Liu, Y., Mao, R., Li, F., Xiao, Y., Wang, Y., Hu, Y., Yu, J., & Zhou, J. (2020). Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer. *Nature Communications*, *11*(1), 1236. https://doi.org/10.1038/s41467-020-15027-z

[44] Nguyen, C., Wang, Y., & Nguyen, H. N. (2013). Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic.

[45] Sheikhpour, R., Sarram, M. A., & Sheikhpour, R. (2016). Particle swarm optimization for bandwidth determination and feature selection of kernel density estimation based classifiers in diagnosis of breast cancer. *Applied Soft Computing*, *40*, 113-131.

[46] Bhardwaj, H., Sakalle, A., Tiwari, A., Verma, M., & Bhardwaj, A. (2018, November). Breast cancer diagnosis using simultaneous feature selection and classification: a genetic programming approach. In *2018 IEEE symposium series on computational intelligence (SSCI)* (pp. 2186-2192). IEEE.

[47] Khuriwal, N., & Mishra, N. (2018, November). Breast cancer detection from histopathological images using deep learning. In *2018 3rd international conference and workshops on recent advances and innovations in engineering (ICRAIE)* (pp. 1-4). IEEE.

[48] Jannesari, M., Habibzadeh, M., Aboulkheyr, H., Khosravi, P., Elemento, O., Totonchi, M., & Hajirasouliha, I. (2018, December). Breast cancer histopathological image classification: a deep learning approach. In *2018 IEEE international conference on bioinformatics and biomedicine (BIBM)* (pp. 2405-2412). IEEE.

[49] Mekha, P., & Teeyasuksaet, N. (2019, January). Deep learning algorithms for predicting breast cancer based on tumor cells. In *2019 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT-NCON)* (pp. 343-346). IEEE.

[50] Yang, X., Wu, L., Ye, W., Zhao, K., Wang, Y., Liu, W., ... & Liang, C. (2020). Deep learning signature based on staging CT for preoperative prediction of sentinel lymph node metastasis in breast cancer. *Academic Radiology*, *27*(9), 1226-1233.

[51] Sha, Z., Hu, L., & Rouyendegh, B. D. (2020). Deep learning and optimization algorithms for automatic breast cancer detection. *International Journal of Imaging Systems and Technology*, *30*(2), 495-506.

[52] Zheng, J., Lin, D., Gao, Z., Wang, S., He, M., & Fan, J. (2020). Deep learning assisted efficient AdaBoost algorithm for breast cancer detection and early diagnosis. *IEEE Access*, *8*, 96946-96954.

[53] Shen, T., Wang, J., Gou, C., & Wang, F. Y. (2020). Hierarchical fused model with deep learning and type-2 fuzzy learning for breast cancer diagnosis. *IEEE Transactions on Fuzzy Systems*, *28*(12), 3204-3218.

[54] Mishra, A. K., Roy, P., & Bandyopadhyay, S. (2021). Binary particle swarm optimization based feature selection (bpso-fs) for improving breast cancer prediction. In *Proceedings of International Conference on Artificial Intelligence and Applications: ICAIA 2020* (pp. 373-384). Springer Singapore.

[55] Jiang, J., Peng, J., Hu, C., Jian, W., Wang, X., & Liu, W. (2022). Breast cancer detection and classification in mammogram using a three-stage deep learning framework based on PAA algorithm. *Artificial Intelligence in Medicine*, *134*, 102419.