International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

# Machine Learning and Deep Learning Techniques for Classification of Breast Cancer: A Survey

**Karthikeya Mallelli[1], Bharath Kumar Nangunuri[2], Sathupati Vaishnavi[3], Greshma Neredimelli[4], Talupula Vaishnavi[5]**

[1,3,4,5]Student, Department of EEE, Vardhaman College of Engineering, Shamshabad, Hyderabad
[2]Student, Department of CSE (AI & ML), Vardhaman College of Engineering, Shamshabad, Hyderabad

**Abstract**

One of the main causes of death is breast cancer or BC. With this survey paper, people can understand what is breast cancer, how it causes death, and what treatments can be used to diagnose breast cancer. Accurate categorization of breast cancer data is essential for cancer diagnosis, and the ability to distinguish benign from malignant tumors can save patients unnecessary medical interventions. The optimal course of action can also be determined using the classification of breast cancer. One well-known area of medical research is the classification of patient populations into benign and malignant conditions. Because machine learning can extract significant characteristics from a collection of medical data, it is widely used in the prediction of breast cancer. Machine learning methods in medical diagnostics have drawn increased interest lately. These technologies, along with the many deep learning approaches that have surfaced recently, allow information based solely on gene expression to be used to inform healthcare decisions, offering insight into sound and suitable healthcare decisions. Tissue texture and breast density are commonly used by doctors and automated technologies as indicators of sickness in diagnostic imaging. Customized screening and preventive decisions can be guided by the exact identification of cancer risk. So, by analyzing this paper people can understand how breast cancer can be detected using machine learning and deep learning techniques and what these techniques make special rather than manual identifying of breast cancer, also it includes different research paper surveys that many professional authors have done. Convolutional neural networks (CNNs) are being used in deep learning to tackle a range of classification and prediction tasks, including breast imaging like Long Short Term Memory(LSTM) and Recurrent Neural Network(RNN) networks.

**Keywords:** Deep Learning; Convolutional Neural Network(CNN);Long Short Term Memory(LSTM); Recurrent Neural Network(RNN); Classification; Diagnosis; Imaging

## 1. INTRODUCTION

A sickness referred to as cancer takes place when any individual’s cells multiply out of range and assault different good-sized locales. The human body contains some trillions of different cells, most cancers can begin at any place. Human cells frequently partition to shape modern cells as required with the aid of the frame using growing and increasing. Cutting-edge cells supplant ancient ones after they kick the bucket

because of maturing or harm. This controlled element can sometimes glitch, causing harmed or atypical cells to multiply and grow once they should not. Tumors are knots of tissue that can be shaped through those cells. Cancerous or kind tumors can both show up. Threatening malignant tumors can be metastasized, or even spread into, organs and also neighboring tissues. This management lets the tumors change into the current region interior of the frame. Threatening tumors are some other period for cancerous tumors. Blood malignancies, counting leukemias, frequently do not create robust tumors, even though numerous cancers do. Kind tumors that don't enter or spread to neighboring tissues are benign. Kind tumors do not often increase again after evacuation, whereas threatening tumors on occasion do. Be that as it may, beneficent tumors can from time to time expand to be or maybe enormous. A few, like generous brain tumors, are likely lethal or motive critical facet results.
Fig 1 depicts the difference between normal cells and cancer cells. Cancer cells are those which grow excessively in the human body.

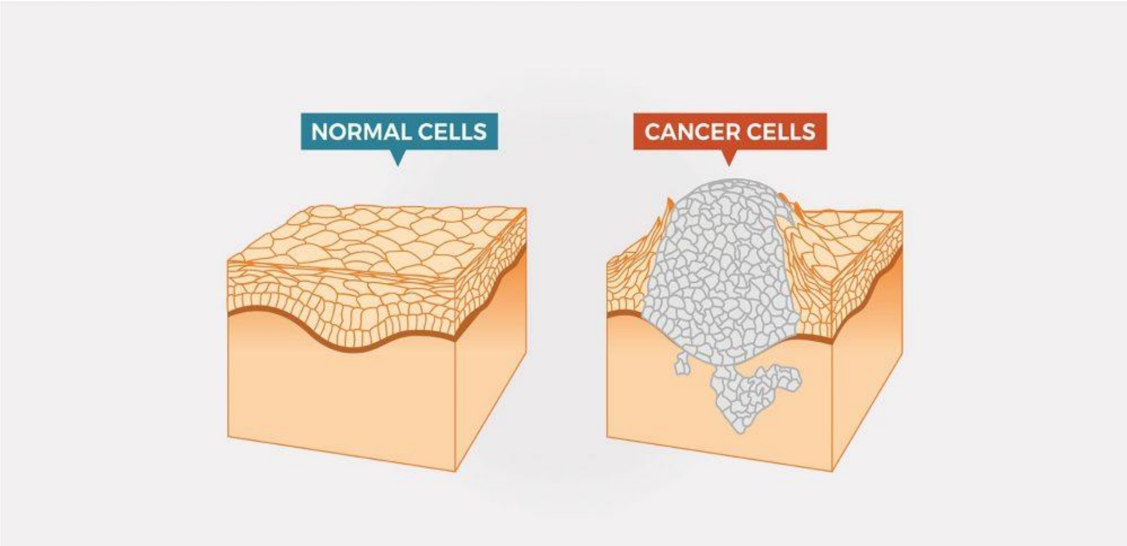


**Fig.1.Cancer Cells[1]**

Fig 2 describes the stages of cancer where the first part shows that the cells are grown normally with the required amount of cells the second part is hyperplasia where the organization is not disturbed but extra cells are grown and the last part is where the organization and level also changed.

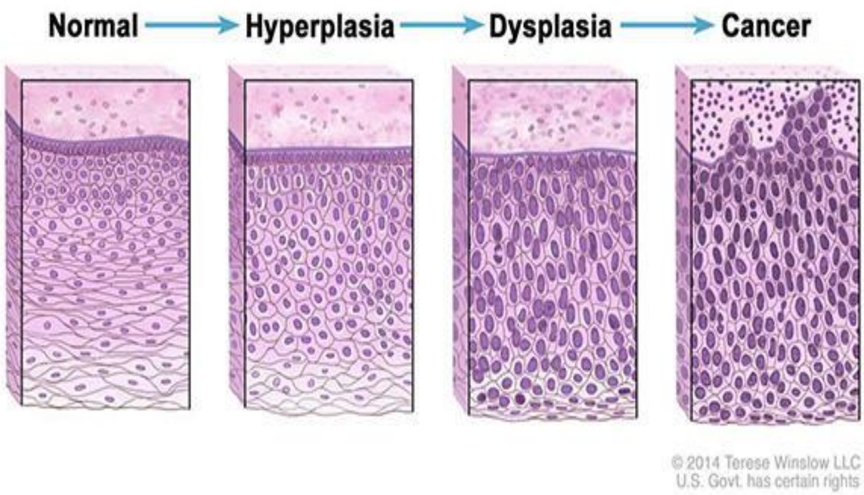


**Fig.2.Stages of Cancer[1]**

### 1.1. Breast Cancer

The illness called breast cancer is a result of aberrant breast cells that proliferate and turn into tumors. Tumors can develop at some point in the body and turn out to be deadly if overlooked. The milk ducts and/or the breast's milk-producing lobules are where breast cancer cells first proliferate. There may be no risk to lifestyles from the early form (in situ). Cancer cells can invade neighboring breast tissue. Tumors produced by this bring about thickening or lumps. Metastasis is the system via which invasive tumors circulate to neighboring lymph nodes or different organs. You can die from metastasis. Treatment is determined by employing the affected person, the most cancer types, and the extent of dissemination. Remedy, radiation remedy, and surgical treatment are all utilized in remedy. Adenocarcinoma is a cancer that forms in epithelial cells that produce fluids or mucus. Tissues with this type of epithelial cell are sometimes called glandular tissues. Most cancers of the breast, colon, and prostate are adenocarcinomas.

### 1.2. Types of Breast Cancer

1. Ductal Carcinoma In Situ (DCIS): Pre-invasive or non-invasive breast cancer is known as DCIS[6]. This indicates that although the cancerous cells lining the ducts have mutated, they have not penetrated the duct walls and reached the breast tissue nearby.
2. Invasive Breast Cancer (IDC/ILC) [6]: Invasive breast cancers are those that have extended into the surrounding breast tissue. Although invasive breast cancer comes in several forms, it accounts for the majority of cases of breast cancer. It contains 2 types.
3. Inflammatory Breast Cancer: It is a form of invasive ductal carcinoma, although there are differences in its symptoms, prognosis, and course of treatment. IBC results in breast inflammation symptoms such as redness and swelling, which are brought on by cancer cells obstructing cutaneous lymph veins and making the breast appear ”inflamed.”
4. Paget Disease of the Breast: An uncommon form of breast cancer that causes damage to the areola (the black circle surrounding the nipple) and the skin around the nipple is called Paget disease of the breast. Typically, just one breast is impacted by Paget illness. It is typically detected in 80–90 per of cases in conjunction with either infiltrating ductal carcinoma.

### 1.3. Stages of breast cancer

1. Stage 0: stage 0 breast cancer is a type where the tumor is just found in the breast cells but not in the surrounding tissues.
2. Stage I: stage 1 is a type where the lymph nodes are not involved and the size of the tumor is much smaller which is about 2cm or even smaller.
3. Stage II: stage 2 is a type where the lymph nodes are involved and the size of the tumor is average which is about 2cm to 5cm.
4. Stage III: The tumor has grown larger, may have invaded nearby tissues, and includes several lymph nodes.
5. Stage IV: The cancer has progressed to remote organs such as the brain, liver, lungs, or bones. The prognosis and therapy for breast cancer vary depending on the disease's kind, stage, and other personal characteristics. Many people with breast cancer now have far better outcomes thanks to advancements in treatment and early detection.

Fig 3 represents the different stages of breast cancer. It contains 5 different stages where the starting stage is the normal stage and the end stage or 4th stage is the one where the tumor spreads throughout the body which is basically known as a malignant type of tumor.

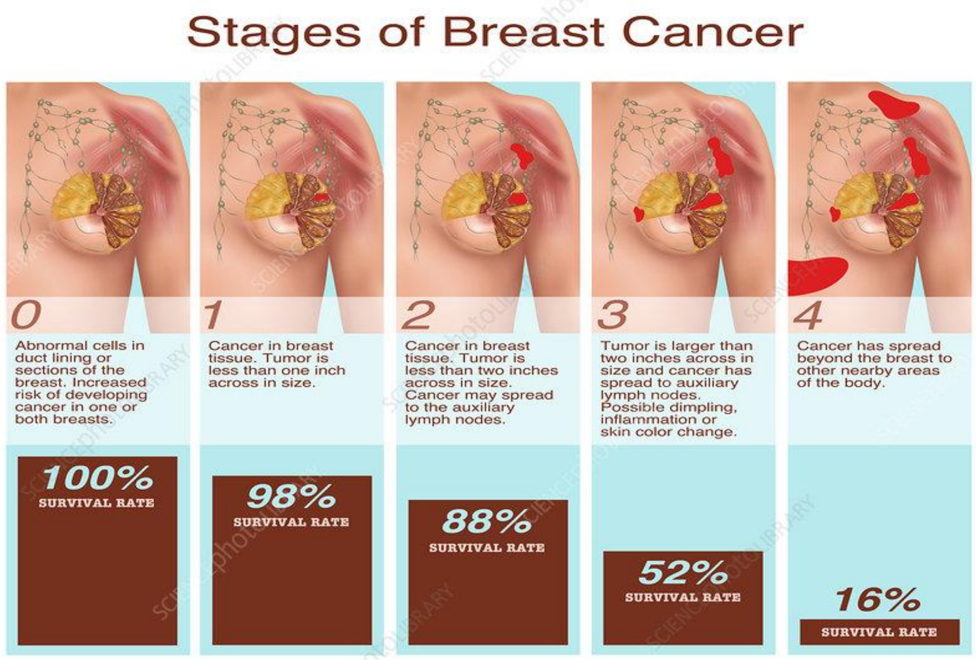


**Fig.3. Stages of Breast Cancer[5]**

### 1.4. Treatment of Breast Cancer

Treatment options depend on the type and stage of breast cancer and may include:

1. Operation: Removal of the whole breast or only the tumor(mastectomy) [3].
2. Radiation Technique: Using high-energy rays to target and kill cancer cells.
3. Chemotherapy: Administration of drugs to destroy cancer cells.
4. Hormone Therapy: Blocking hormones that fuel some types of breast cancer.
5. Targeted Therapy: Targeting specific molecules involved in cancer growth.

### 1.5. Overview of Machine Learning and Deep Learning for Cancer

Cancer continues to be one of the most difficult medical problems to solve. Because of its complexity and diversity, fast detection and precise prediction are mandatory for successful therapy. To fight this, machine learning (ML) techniques and deep learning (DL) techniques have become formidable weapons that provide cutting-edge approaches to cancer prediction and better patient outcomes. Machine Learning for Cancer Prediction: The creation of algorithms for machine learning, a branch of artificial intelligence, enables computers to recognize patterns in data without the need for explicit programming. In the context of cancer prediction, machine learning (ML) makes use of enormous datasets, such as genetic data, patient records, and medical imaging, to find minute patterns suggestive of the onset of cancer. Cancer prediction models use a variety of machine learning (ML) techniques, including ensemble methods, supervised learning, and unsupervised learning. Deep neural networks, which are a subset of machine learning, are the subject of deep learning. Healthcare is one of the many hard jobs that these networks are well-suited for since they can automatically extract nuanced features from raw data. DL has demonstrated impressive efficacy in cancer prediction in domains including radiography and pathology. DL is a master at using methods like convolutional neural networks (CNNs) are used for image-based cancer prediction. For instance, DL algorithms may examine mammograms for breast cancer screening and spot minute anomalies that might go unnoticed by humans. The precision of cancer diagnosis and prognosis is increased by the capacity of deep learning (DL) models to learn hierarchical representations of features, which allows them to identify subtle patterns.

In Section **2** of the study, the relevance of machine learning and deep learning in the context of breast cancer prediction is further highlighted. Section **3** is devoted to a literature review of studies written by different authors. The lesson learned from this survey paper is presented in the concluding and last Section **4**, Conclusion.

## 2. MACHINE LEARNING AND DEEP LEARNING IN THE CLASSIFICATION OF BREAST CANCER

### 2.1. Machine Learning

Definition: Machine learning is the branch of artificial intelligence that deals with developing statistical models and techniques that enable computers to do tasks without the need for explicit programming. The basic idea is to empower machines to learn from data and progressively improve their performance. Features: Machine learning algorithms use patterns they've discovered in past historical data for forecasting and also to decide on previously unseen data. Machine learning comes in different flavors: reinforcement learning, unsupervised learning, and supervised learning

**1) Supervised Learning:** Supervised learning is defined as training an algorithm using a labeled dataset comprised of input-output pairs. In this paradigm, the algorithm generalizes patterns from the given instances to learn how to transfer input data to appropriate output labels. The algorithm is guided by an existing set of goal outputs to train it to provide predictions or classifications on fresh, unknown data. It is classified into two types. They are:

- Classification: The objective of a supervised learning task called classification is to categorize or label input data according to predetermined criteria based on its characteristics. The method guesses the class or category to which new, unseen instances belong after learning from labeled training examples.
- Regression: Similar to classification, regression is a supervised learning technique that entails the prediction of a continuous numerical output. The algorithm in this challenge learns the relationship between input features and a continuous target variable to generate predictions on fresh data.

Fig 4 describes the architecture of supervised learning which includes labels for the training data and then training the model using different algorithms for prediction.

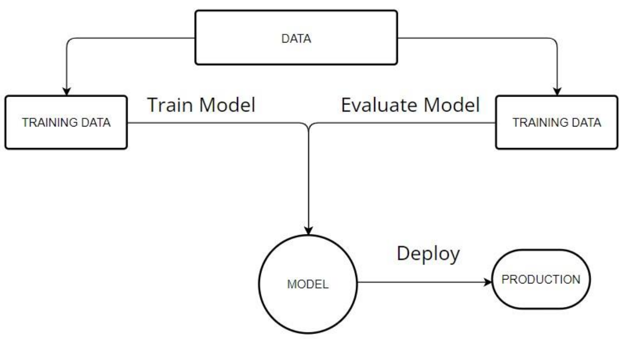


**Fig.4.Supervised Learning**

**2) Unsupervised Learning:** Unsupervised learning refers to training algorithms using unlabelled data without the need for explicit instructions or predetermined results. Unsupervised learning aims to reveal underlying structures, relationships, or patterns in the data without the need for labeled samples to be used as a guide. Unsupervised learning techniques investigate the underlying structure of the data to find similarities, clusters, or underlying features. It is classified into two types. They are:

- Clustering: Clustering is defined as identifying the common patterns based on the unlabelled data that is given as input.
- Association: Association, on the other hand, focuses on identifying relationships or associations between different variables in a dataset. This is generally used in market analysis and recommendation systems. The primary goal is to discover patterns where the occurrence of one event is correlated with the occurrence of another.

**3)** Reinforcement Learning: Reinforcement learning is a method that trains on itself which is self-learning. It contains rewards and penalties. Rewards are given to encourage the model and penalties are given as punishments.

Fig 5 describes the architecture of reinforcement learning which includes the agent and environment. The environment is the one where we do the action and the agent is the one which gains reward or penalty based on the action.

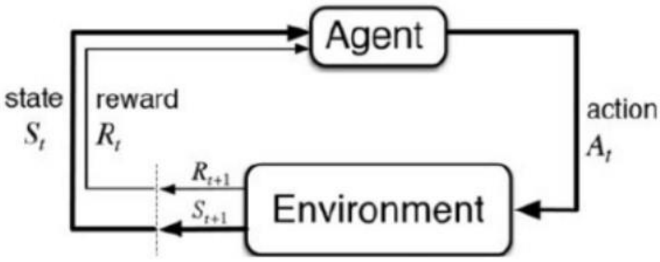


**Fig.5. Reinforcement Learning[14]**

**2.2. Deep Learning**

Deep Learning techniques contain multiple layers to handle large amounts of data. Deep neural networks are trained on vast volumes of data. Characteristics: Artificial neural networks and other deep learning architectures are made up of several layers of networked nodes, or neurons, which can automatically extract and learn intricate characteristics from the input. Natural language processing, picture and audio recognition, and other fields have seen tremendous success with deep learning.

**1) Convolutional Neural Networks (CNNs):** Deep learning relies heavily on Convolutional Neural Networks (CNNs), which are specialized neural networks made for handling and interpreting structured grid data, like photographs. They are particularly good at tasks like feature extraction, object detection, and image recognition. Convolutional layers are a tool that CNNs use to capture hierarchical patterns and features in input data by methodically applying filters or kernels. By identifying basic patterns in the early layers and gradually integrating them to recognize sophisticated structures in the deeper levels, these layers allow the network to learn spatial hierarchies of features. Layers of pooling further reduce the computing needs by downsampling the learned features. CNNs perform exceptionally well in image-related tasks due to their architecture's capacity to autonomously learn hierarchical representations, which has greatly advanced computer vision and pattern identification.

Fig 6 describes the architecture of the convolutional neural network which includes the convolutional layers and pooling layers. Convolutional layers are for feature extraction and pooling layers are for the spacial adjustment

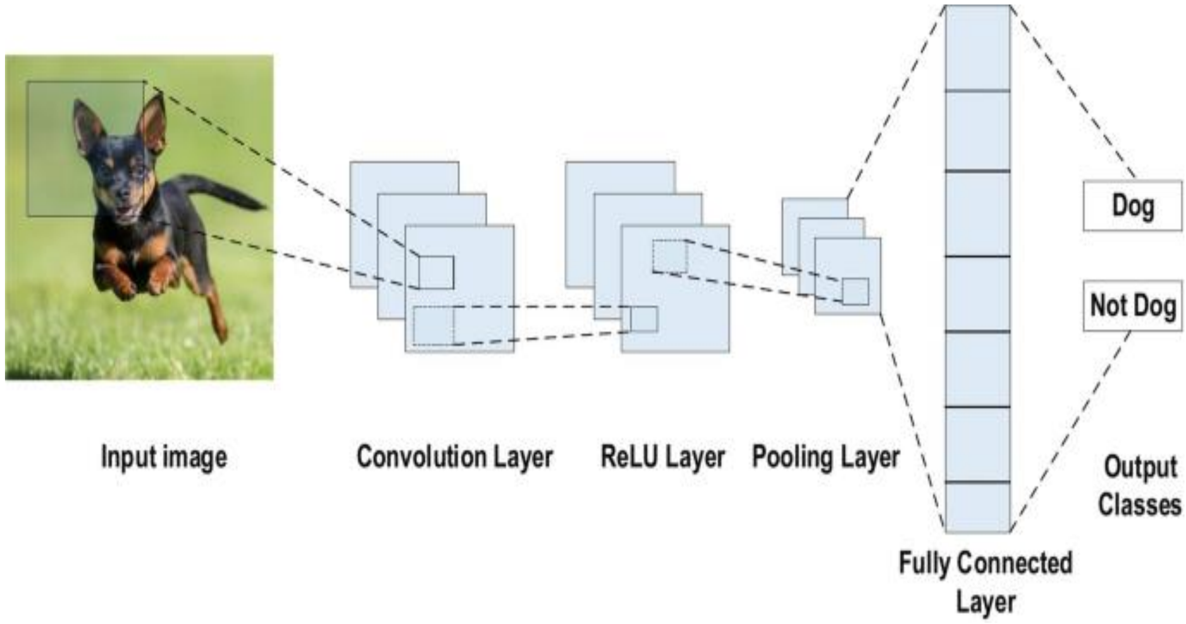


**Fig.6. Convolutional Neural Network[15]**

2) **Autoencoders:** A particular kind of artificial neural network used for unsupervised learning that is tailored for data encoding and decoding tasks is called an autoencoder. With the help of an encoder and a decoder, it seeks to decrease the amount of input data to learn a compact representation of it. The input is compressed or encoded by the encoder, and the original input is reconstructed from this representation by the decoder. Because autoencoders can extract important features and patterns from input data while removing irrelevant details, they are used in a wide range of fields.

3) In deep learning, ensemble learning is the process of merging several separate models to produce a predictive model that is more reliable and accurate. This strategy makes use of the diversity of several models to increase performance and improve generalization. Ensemble approaches in deep learning typically entail training several neural networks with different initializations, designs, or training data subsets, and then combining their predictions. The ensemble algorithms bagging, boosting, and stacking are widely used in deep learning. Several parallel models are constructed using bagging techniques like Random Forests, each of which is trained on a distinct subset of data. Boosting, which is represented by algorithms such as Gradient Boosting and AdaBoost, trains models in a sequential fashion with an emphasis on fixing the mistakes made by their forerunners. To create a final prediction, stacking combines predictions from numerous models using a different model, usually a meta-learner. In a variety of deep learning problems, these ensemble approaches help to increase the robustness and performance of the models.

4) **Transfer Learning:** In the context of deep learning, transfer learning is a method for the machine learning paradigm where a neural network model that has already been trained and optimized for one task is modified or adjusted to address a related but unrelated task. When there is a limited quantity of labeled data available for a particular job, the model can perform better and expedite learning by utilizing information from another domain. This method makes use of the pre-trained model's hierarchical feature representations, which let it recognize and apply broad patterns to a variety of tasks.

5) The yields of the past states are given as input to the current state in repetitive neural systems (RNN). RNN's covered-up layers can recollect information. The yield from the past state is utilized to overhaul the concealed state. Since RNNs have long short-term memory (moreover known as the capacity

to keep in mind past inputs), they can be utilized for time arrangement expectations. The architecture of the recurrent neural network includes different gates like reset gate and update gate.

## 3. LITERATURE REVIEW

In a groundbreaking effort, Akhil Jabbar et al. [8] defined ensembling learning in the prediction of breast cancer. Focusing on the combination of the Bayesian network with the radial basis function, their method achieved an accuracy of 97 percent on the Wisconsin dataset which contains 540 rows and 32 columns which indicate features. Breast Cancer, a significant global threat demands an accurate prediction and timely management.

In. [10] authors have used the BreaKHis dataset which consists of 7909 images of breasts which include 8 types of benign tumors and malignant tumors. The cross-validation technique uses 25 per of the training dataset. To deal with the dataset a 5 CNN layer topology which includes filters of 3*3 and 2 fully connected layers are used. They worked on different pre-trained models to improve the accuracy in classifying the images and among them, they achieved an accuracy of 92 per on ResNet50 which is the highest percentage of accuracy.

In. [11] authors have used the breast histopathology images as a dataset which consists of 7000 images and achieved an accuracy of 95.55 per. For the classification of the breast images, they used convolutional neural networks.

In. [12] authors have used the Mammography images dataset from Kaggle. Healthcare always expects a higher rate of accuracy so they focused on deep learning methods for breast cancer classification and they achieved an accuracy of 97 per with the help of the RFE (recursive feature extraction) technique.

Authors of [13]. has proposed different machine-learning models for the detection of breast cancer. They proposed RF, GB, AB, Bagging, and KNN with different accuracies and they achieved a higher rate of accuracy on RF, GB, and AB because of their capability of combining weaker models and boosting and also generating randomness to the model. The others have a lesser rate of combining the accuracies.

The following Table 1 depicts how various ML/DL Algorithms are used by different researchers for breast cancer detection.

**Table 1 SUMMARY OF MACHINE LEARNING AND DEEP LEARNING MODELS ON BREAST CANCER CLASSIFICATION**

| Model | Reference | Summary | Accuracy |
|---|---|---|---|
| Ensembling Learning of RBF and Bayesian Network | [8] | Ensembling increased accuracy compared when to previous models | An accuracy of 97.42 % was achieved. |
| Hybrid CNN Algorithm | [10] | Computational cost is low. CNN performs better in disease detection fast compared to others. | 95.6% accuracy rate was achieved. |
| CNN | [11] | The existing is computationally expensive and a 95.6% | 99.55%. |

| | | | |
|---|---|---|---|
| | | accuracy rate was achieved. 99.55%. difficult, so deep learning techniques like convolutional neural networks (CNN) are used for the classification of histopathological images. | |
| CNN, Deep Neural Network, and RFE | [12] | Increasing the number of layers in neural networks increases the extraction of features. | 97% |
| KNN, AB, RF and GB | [13] | They proposed RF, GB, AB, Bagging, and KNN with different accuracies and they achieved a higher rate of accuracy on RF, GB, and AB because of their capability of combining weaker models and boosting and also generating randomness to the model. | The GB, RF, and AB models achieved 100% accuracy, the Bagging, KNN, and MLP models achieved a performance of 99.56%, 95.82%, and 96.92%. |

## 4. CONCLUSION AND RECOMMENDATIONS

In conclusion, this literature review examined the wide range of approaches to breast cancer prediction with a variety of deep learning methods and also machine learning methods, mammography and histopathology image analysis, and numerical data from the Wisconsin dataset. The thorough analysis emphasized the critical role that cutting-edge technology plays in enhancing accuracy and efficiency while highlighting the importance of early detection in the treatment of breast cancer. The amalgamation of deep learning and machine learning models showed auspicious outcomes, as multiple investigations illustrated the proficiency of these methodologies in capitalizing on intricate patterns found in medical imaging and numerical attributes for precise forecasting. As essential diagnostic modalities, mammography, and histopathological images were essential for training and assessing models. As far as we could tell, the authors of earlier studies employed a variety of datasets, including numerical and image datasets. The Wisconsin dataset, which is standard data for breast cancer prediction, was selected by the authors for numerical data. They were able to attain improved accuracies to the best of their abilities while boosting and other ensembling learning techniques produced better results by merging weak learners. Additionally, because CNN uses filters and kernels for feature extraction, picture dataset histopathology produces results with improved accuracy. Therefore, we will advise using CNN and SVM algorithms to detect breast cancer using histopathological pictures and Wisconsin datasets.

**REFERENCES**

1. DocterFly Study https://doctorfly.co.uk/6-differences-between-normal-and-cancer-cells/ (Last accessed: 25/05/2024)
2. Mayo Clinic, Symptoms https://www.mayoclinic.org/diseases and Causes conditions/breast cancer/symptoms-causes/syc-20352470 (Last accessed: 25/05/2024)
3. World Health Organization, breast-cancer https://www.who.int/news=room/fact-sheets/detail/breast-cancer (Last accessed: 25/05/2024)
4. Cureus, mammography https://www.cureus.com/articles/124192 correlation-of-mammography-ultrasound-and-sonoelastographic f indings-with-histopathological-diagnosis-in-breast-lesions!/(Last accessed: 25/05/2024)
5. Science Photo Library, Stages of Breast https://www.sciencephoto.com/media/1137060/view/breast-cancer stages-illustration (Last accessed: 25/05/2024)
6. Cancer, breast-cancer https://www.cancer.org/cancer/types/breast cancer/about/types-of-breast-cancer.html (Last accessed: 25/05/2024)
7. National Breast Cancer, anatomy https://nbcf.org.au/about-breast cancer/diagnosis/breast-cancer-anatomy/ (Last accessed: 25/05/2024)
8. Akhil, Jabbar. (2021). Breast Cancer Data Classification Using Ensemble Machine Learning. KKU ENGINEERING JOURNAL. 48. 65-72.
9. S. J. Malebary and A. Hashmi, ”Automated Breast Mass Classification System Using Deep Learning and Ensemble Learning in Digital Mammogram,” in IEEE Access, vol. 9, pp. 55312-55328, 2021, doi: 10.1109/ACCESS.2021.307129
10. Masud M, Eldin Rashed AE, Hossain MS. Convolutional neural network-based models for diagnosis of breast cancer. Neural Com put Appl. 2022;34(14):11383-11394. doi: 10.1007/s00521-020-05394-5. Epub 2020 Oct 9. PMID: 33052172; PMCID: PMC7545025.
11. F. Shahidi, S. Mohd Daud, H. Abas, N. A. Ahmad and N. Maarop, ”Breast Cancer Classification Using Deep Learning Approaches and Histopathology Image: A Comparison Study,” in IEEE Access, vol. 8, pp. 187531-187552, 2020, doi: 10.1109/ACCESS.2020.3029881.
12. V. Asha, B. Saju, S. Mathew, A. M. V, Y. Swapna and S. P. Sreeja, ”Breast Cancer classification using Neural networks,” 2023 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE), Bengaluru, India, 2023, pp. 900 905, doi: 10.1109/IITCEE57236.2023.10091020.
13. Iparraguirre-Villanueva, Orlando, Andr´es Epifan´ıa-Huerta, Carmen TorresCecl´en, John Ruiz-Alvarado, and Michael Cabanillas-Carbonel. ”Breast cancer prediction using machine learning models.” (2023).
14. Mahesh, Batta. (2019). Machine Learning Algorithms-A Review. 10.21275/ART20203995.
15. Alzubaidi, L., Zhang, J., Humaidi, A.J. et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data 8, 53 (2021).
16. Sarker, I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN COMPUT. SCI. 2, 160 (2021)