



# Journal of Innovations in Applied Pharmaceutical Science [JIAPS]

Content available at: [www.saap.org.in](http://www.saap.org.in) ISSN: 2455-5177



## APPLICATION OF MACHINE LEARNING TECHNIQUES TO IMPROVE DETECTION, DIAGNOSIS & PREDICTION OF BREAST CANCER: A COMPARATIVE ANALYSIS

Shobhit Shrotriya<sup>1</sup>, Nizar Banu P K<sup>2</sup>, Avi Kulkarni<sup>3</sup>

<sup>1</sup> PhD Scholar, Department of Computer Science, CHRIST (Deemed to be University), Bangalore – 29

<sup>2</sup> Associate Professor, Department of Computer Science, CHRIST (Deemed to be University), Bangalore – 29

<sup>3</sup> President, Syneos Health, California, USA

### Article History

Received: 04-10-2023

Revised: 22-10-2023

Accepted: 19-11-2023

**Keywords:** machine learning (ML), deep learning (DL), breast cancer prediction, datasets.



### Abstract

There is an increasing cancer burden in India across ages and sexes. The most significant cancer incident rate in females is 'Breast Cancer'. Early detection and treatment are the key to lower mortality rate and better survival rates for cancer patients in the country. This review paper provides an understanding of the various types of breast cancers, their symptoms, causes, current detection and diagnosis methods. The paper presents different Machine Learning (ML) techniques that are in development for the detection and diagnosis of breast cancer. The objective of the paper is to highlight outcomes of some select previous studies between 2016 to 2020 using various ML techniques and summarize the selected algorithms which can be used for breast cancer prediction and diagnosis. In our paper, we have also made an attempt to implement the Convolutional Neural Networks (CNN) model on the Breast Cancer Wisconsin (Diagnostic) dataset, whose results are presented and discussed.

This article is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License. Copyright © 2023 Author[s] retain the copyright of this article.



### \*Corresponding Author

Shobhit Shrotriya

<https://doi.org/10.37022/jiaps.v8i3.519>

### Production and Hosted by

[www.saap.org.in](http://www.saap.org.in)

### Introduction

Accounting for 14% of all cancers in Indian women and amongst the leading causes of death, breast cancer is the most common type of cancer. India also has far lower survival rates of 66.1% for breast cancer compared to 90.2% in the USA. Almost 50% of all cases are in the age group of 25-50 years and more than 70% of the cases present in the advanced stage with poor survival and high mortality [15, 16, 17]. The report in [17] says, every four minutes an Indian woman is diagnosed with breast cancer. Women from their early thirties to fifties are at considerable risk of developing breast cancer, and the incidence risk increases till its peak by the time they reach 50-64 years of age [17]. One in twenty-eight Indian women is likely to develop breast cancer during her lifetime and the survival rates in India are low because it is not detected at the initial stage. This can be reduced by increasing awareness of how it can be detected and diagnosed early. There are several machine learning & deep learning algorithms that are being used by researchers and clinicians for the prediction of breast

cancer, however, finding the most suitable one which can provide sustainable and accurate prediction is imperative to ensure early detection and diagnosis.

Understanding breast cancer is important [18, 19, 20]. Breast cancer originates in the breast tissue; occurs when breast cells mutate and grow out of control, creating a mass of tissue (tumor). Similar to any other cancers, breast cancer can invade and grow into the tissue surrounding the breast and may also travel to other parts of the body and form new tumors. Such a phenomenon called metastasis. There are several different types of breast cancer, including:

1. **Infiltrating (invasive) ductal carcinoma (IDC)**- Starts in the milk ducts of the breast and breaks through the wall of the duct and spreads to surrounding breast tissue. Making up about 80% of all cases, this is the most common type of breast cancer [20].
2. **Ductal carcinoma in situ (DCIS)**- Also termed as Stage 0 breast cancer, ductal carcinoma in situ is considered by some to be precancerous because the cells have not spread beyond the milk ducts. The condition is very treatable but significant care is necessary to prevent the cancer from becoming invasive and spreading to other tissues [20].
3. **Infiltrating (invasive) lobular carcinoma (ILC)** - This cancer forms in the lobules of the breast and spreads to surrounding breast tissue. It accounts for 10% to 15% of breast cancers [20].

4. **Triple negative breast cancer (TNBC)** - Making up about 15% of all cases, triple negative breast cancer is one of the most challenging breast cancers to treat because it does not have three of the markers associated with other types of breast cancer. This makes prognosis and treatment difficult [20].
5. **Inflammatory breast cancer**- This cancer is very rare but aggressive and resembles an infection. Women with inflammatory breast cancer usually notice redness, swelling, pitting, and dimpling of their breast skin. It is caused by obstructive cancer cells in their skin's lymph vessels [20].
6. **Paget's disease of the breast** - This cancer affects the nipple's skin and areola [20].
7. **Angiosarcoma** - These are rare and constitute approximately 1% of all different types of breast cancers. They initiate in the cells that line blood vessels or lymph vessels and impact the breast tissue or the skin of the breast. Prior radiation therapy in that area may sometimes be the reason of their origination [20].
8. **Phyllodes tumor** - These tumors are very rare. They develop in the connective tissue of the breast, called stroma of the breast. These are mostly benign, but some of them may be malignant (cancer).

Diagnosis of breast cancer can be done using various techniques.

**Mammogram:** A mammogram is used in breast cancer prevention. These are special X-ray images that detect changes or abnormal growths in the breast.

**Ultrasonography:** sound waves are used to capture pictures of the tissues inside of the breast which helps to diagnose breast lumps or abnormalities.

**Positron emission tomography (PET) scanning:** Special dyes are used to highlight suspicious areas. The healthcare provider injects a special dye into the veins to capture images with scanner.

**Magnetic resonance imaging (MRI):** Magnets and radio waves are used to produce clear and detailed images of the structures inside of the breast.

Machine learning and deep learning algorithms help us in diagnoses, detection, and prediction of such types of breast cancer discussed above. The associated data mining techniques enable us to gather meaningful information about the breast cancer patients.

#### ML & DL algorithms for breast cancer detection, diagnosis & prediction

##### Understanding ML and DL concepts

Machine learning is a branch of artificial intelligence combined with computer science that focuses on the use of data and algorithms to mimic the way that humans learn [11]. Figure 1 shows the world of data science [22].

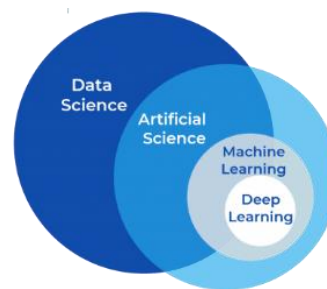


Fig. 1: The world of Data Science

Machine learning algorithms are grouped into four buckets [12]. Figure 2 presents the types of machine learning algorithms.

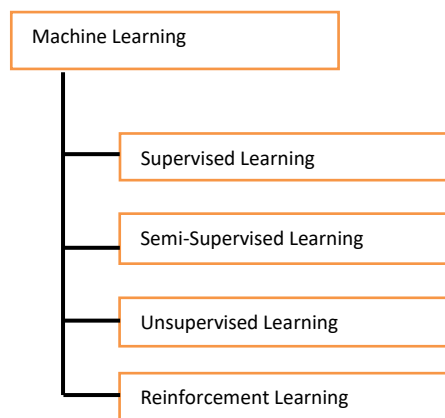


Fig. 2: Types of ML Algorithms

**(a) Supervised learning:** These algorithms use labeled data to predict the output from the input data.

Supervised learning can be further categorized as **Classification Model and Regression Model**. Regression techniques are used to predict continuous values, while classification algorithms predict discrete. Examples of classification algorithms include Decision Tree, Naïve Bayes, Random Forest, Logistic Regression and K-nearest neighbor (KNN).

**(b) Unsupervised learning:** These algorithms are trained using unlabeled data, which is easy to collect and store.

Unsupervised learning is broken down into two subcategories: **Clustering, Association and Dimensionality reduction**. Clustering is a data mining technique for grouping unlabeled data based on their similarities or differences. K-means is one of the algorithms which is a part of clustering technique. Association uses different rules to find relationships between variables in a given dataset. Dimensionality reduction is a learning technique used when the number of features (or dimensions) in a given dataset is too high [13].

**(c) Semi-supervised learning:** It lies between supervised learning and unsupervised learning. Machines are trained using both labeled and unlabeled data.

**(d) Reinforcement learning:** This uses observation gathered from the interaction with the environment to act in a way that maximizes the reward or minimizes the risk. As an algorithm, it continuously studies its environment until it explores all possibilities [12].

The four most commonly used ML algorithms include:

1. K-Nearest Neighbor
2. Linear Regression
3. Decision Tree
4. Naïve Bayes

Given the fact that there are multiple algorithms to choose from, but there is not one which suits every situation, some of the critical questions to ponder before deciding which algorithm to use are as follows:

1. What is the size of the data that you will be working with?
2. What is the type of data that you will be working with?
3. What kind of insights are you looking from the data?
4. How will those insights be used?

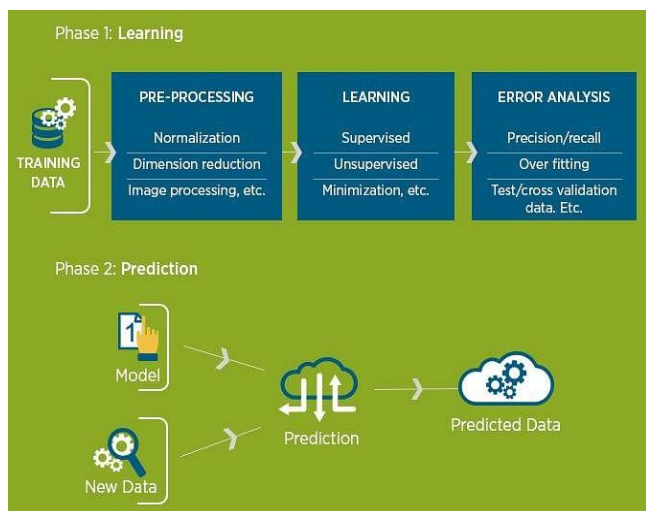


Figure 3 shows a simple process flow depicting how ML works.

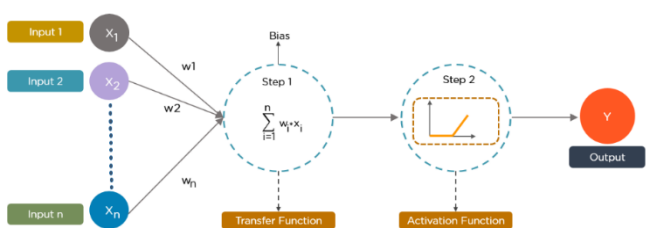
**Fig. 3: The Machine learning process flow**

(Source:[https://www.simplilearn.com/ice9/article\\_detailed\\_content\\_img/machine-learning-process.jpeg](https://www.simplilearn.com/ice9/article_detailed_content_img/machine-learning-process.jpeg)). The concepts of deep learning and machine learning tend to be used interchangeably. It is important to note that they are both sub-fields of artificial intelligence [11]. In Deep Learning, the unstructured data can be ingested in its raw form (e.g. text, images), and it can automatically determine the set of features which distinguish different categories of data from one another. Unlike machine learning, it does not require human intervention to process the data, thereby enabling the use of larger datasets.

Deep Learning uses artificial neural networks to perform sophisticated computations on large amounts of data. It is a type of machine learning that works based on the structure and function of the human brain [14].

A neural network is structured like the human brain and consists of artificial neurons, also known as nodes. These nodes are stacked next to each other in three layers:

- The input layer
- The hidden layer(s)
- The output layer



**Fig.4: A fundamental 'Neural Network' construct**

(Source:[https://www.simplilearn.com/ice9/free\\_resources\\_article\\_thumb/ArtificialNeuralNetwork.PNG](https://www.simplilearn.com/ice9/free_resources_article_thumb/ArtificialNeuralNetwork.PNG))

Data provides each node with information in the form of inputs. The node multiplies the inputs with random weights, calculates them, and adds a bias. Finally, nonlinear functions, also known as activation functions, are applied to determine which neuron to fire. The deep learning algorithms depend upon ANNs (Artificial Neural Networks) that mirror the way the brain processes information. The algorithms while in the training process, use unknown elements in the input distribution to extract features, group objects, and identify useful data patterns. This approach enables building efficient and effective models.

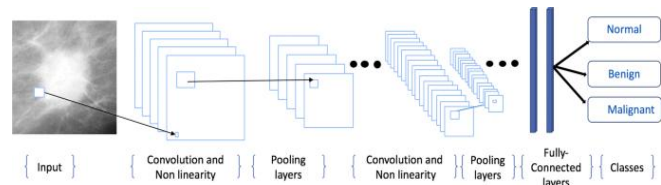
There are multiple types of algorithms used in DL. The top 10 most popular ones are listed below.

1. CNN- Convolutional Neural Networks
2. LSTM - Long Short-Term Memory Networks
3. RNN - Recurrent Neural Networks
4. GAN - Generative Adversarial Networks
5. RBFN - Radial Basis Function Networks
6. MLP - Multilayer Perceptron's
7. SOM - Self-Organizing Maps
8. DBN - Deep Belief Networks
9. RBM - Restricted Boltzmann Machines
10. Auto encoders

Our focus below is on CNN which is the most pre-dominantly used DL technique in the detection and prediction of cancers.

**Convolutional Neural Networks (CNNs)**

CNN's, also known as ConvNets, consist of multiple layers and are mainly used for image processing and object detection. CNN's are widely used to identify satellite images, process medical images, forecast time series, and detect anomalies. CNN can take an image as an input, then assigns weights and related biases to varied aspects in the image to enable differentiation amongst images. CNN has three main layers: Convolutional layer, Pooling layer and Fully Connected layer. The convolution layer that has several filters to perform the convolution operation. They have a Rectified Linear Unit (ReLU) layer to perform operations on elements. The output is a rectified feature map. In the convolution layer, the output is a grid array huge in size. To reduce the size of this array, we use an algorithm max pooling for down-sampling. The pooling layer then converts the resulting two-dimensional arrays from the pooled feature map into a single, long, continuous, linear vector by flattening it. A fully connected layer forms when the flattened matrix from the pooling layer is fed as an input, which classifies and identifies the images. Figure 5 presents a sample CNN architecture.



**Fig.5: A sample CNN architecture**

(Source:<https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-019-2823-4>)

There are various machine learning algorithms and methods that could be used for diagnosing breast cancer, such as Support Vector Machine (SVM), Random Forest, K-Nearest Neighbor, etc. Even though support vector machines are widely used for different tasks, and have shown good results, CNNs are showing a better performance because of their ability to

extract important features and then apply Transfer Learning that helps to reuse the previous models for a running task and improves the used model architecture to reach higher accuracy and precision.

### Fundamentals of developing a ML/DL model

A ML/DL model development starts with the basic step of data preparation. Data preparation involves collection of data from hospitals or the open-sourcedata repositories such as UCI, Kaggle available on the internet. Once the datasets are secured, they can be combined in order to have a larger dataset, which

provide improved accuracy to the results. An important sub-step in data preparation phase is data visualization. This sub-step involves visualizing the data in the form of correlation graph, scatter plot graph and histograms. This helps in getting a better idea about the dataset.

Once the data preparation is completed, it moves to the pre-processing stage. Pre-processing happens in three sub-steps - In the first sub-step, data is transformed and scaled to a certain range, to enable the model in ease of results prediction. The next sub-step called data cleaning; unwanted data is removed from the dataset. The last sub-step is data anonymization, which involves removing of personal details such as name, ID, etc. of the patient from the dataset.

Pre-processing of data is followed by dataset parting. Dataset parting includes parting the dataset into training set and testing set. Typically, 80% of information is taken for training as the model requires larger datasets to learn and predict with improved accuracy. The remaining 20% of the information is taken for testing. Once the data parting is done, the model is trained with the help of an efficient algorithm. Once the model is trained, it is evaluated/tested through the 20% of the data that was secured for testing purposes. The testing dataset produces the percentage of accuracy shown by the model. The last step is the validation phase. K fold cross validation is the technique used for validation. This technique involves splitting the dataset into k groups. Suppose if k is taken as 10, then the dataset is split into 10 groups. One group is taken for training and the remaining groups are taken for testing. This process is repeated until all k groups are taken for testing. Upon completion of cross validation, the model predicts if thepatient has benign tumor or a Malignant tumor. Figure 6 shows the workflow for model development.

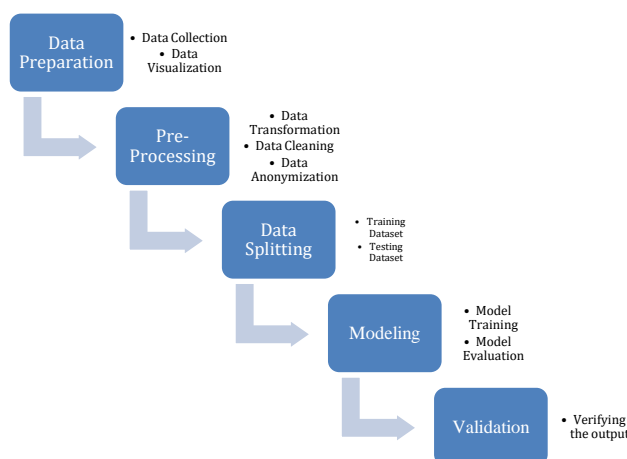


Fig. 6: Workflow for Model development

### Comparative review of ML techniques for breast cancer prediction

A simple search on Google Scholar using the key words such as research papers, Machine Learning, Deep Learning, Breast Cancer prediction yielded 17,800 results for the time range 2016 to 2022. These research papers from the well-known industry platforms such as researchgate.com, sciencedirect.com, iee.org, nature.com, nih.gov, etc. to name a few.

For breast cancer detection and prediction through the application of ML/DL techniques, the major challenge has always been the availability of datasets. Every algorithm requires a large amount of training data for its computational measurements. Also, at times it is observed that either only the numeric dataset is available, or image datasets are available. Availability of both datasets for the same set of cancer patients is usually rare, however many researchers are now putting their efforts to get the datasets of cancer patients in the form of medical images, which may contain the confidential information about the cancer patients. Many of these datasets are open source and available in the form of raw images. To handle the issue of limited dataset availability, the researchers are now using the data augmentation schemes, that consist of some key features including cropping, filtering, rotating, cleaning etc.

## Discussion

### Comparative analysis of ML/DL algorithms for select research papers

Our main aim was to focus on papers that included prediction of breast cancer using machine learning and deep learning techniques. We randomly picked up 10 papers within the time range of 2016 – 2020. We reviewed each of the thoroughly to understand the different ML/DL techniques that were applied for the breast cancer detection and prediction. The datasets were a combination of numerical, categorical and image data. The Table 1 provides a comparative summary of the ML/DL techniques for breast cancer prediction. As is evident from the Table 1, there are different methods of machine learning and deep learning that are currently being used for cancer prediction. Different datasets produce different accuracies with these algorithms.

Table 1: Comparative summary of the ML/DL techniques for breast cancer prediction

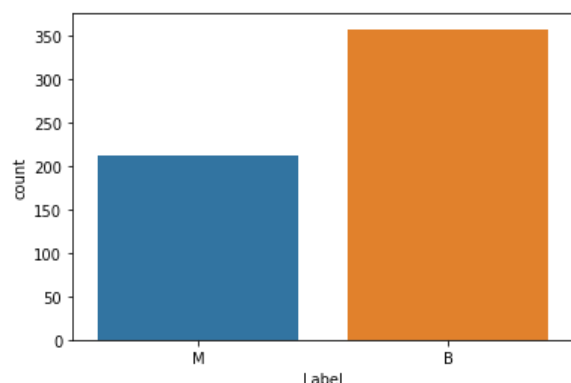
Dataset	Data Type / Attributes	Algorithm	Accuracy	Year	R#
Breast Cancer Wisconsin (Diagnostic) Dataset	Numeric/32	NB SVM ENSEMBLE	97.3% 98.5% 97.3%	2016	[1]
Wisconsin Breast Cancer Dataset at UCI ML repository	Numeric/35	NB C4.5 SVM	67.17% 73.73% 75.75%	2016	[2]
Wisconsin Breast Cancer (Original) Dataset	Numeric/10	SVM C4.5 NB KNN	97.13% 95.13% 95.99% 95.27%	2016	[3]
Wisconsin Breast Cancer Dataset at UCI ML repository	Numeric/11	J48 NB SMO	99.24% 99.12% 99.56%	2020	[4]
Breast Cancer Dataset at UCI ML repository	Categorical/10	J48 NB SMO	98.2% 76.61% 95.32%		
Breast Cancer Dataset at UCI ML repository	Categorical/11	J48 NB SVM GRNN	91% 89% 89% 91%	2019	[5]
www.bioimage.ucsb.edu	Image/16	MLP LMT Random Forest Rotation Forest NB SMO J Rip PART	80% 82.9% 80% 85.7% 82.9% 85.7% 82.9% 77.1%	2017	[6]
BreakHis breast image dataset <a href="http://web.inf.ufpr.br/vri/breast-cancer-database">http://web.inf.ufpr.br/vri/breast-cancer-database</a>	Image	DNN (CNN+LSTM) 40x 100x 200x 400x	90% 90% 91% 90%	2018	[7]
microscopy BACH 2018 grand challenge dataset (ICIAI 2018)	400 histopathological images	Deep CNN	96.10%	2020	[8]
Breast Cancer Coimbra dataset at UCI ML repository	Numeric/10	NB Decision Tree KNN Random Forest	59% 70% 48% 70%	2019	[9]
Kaggle dataset for Breast Histopathology images	Image	NB Decision Tree KNN Random Forest	84% 87% 88% 92%		
UC Irvine Machine Learning Repository	Categorical/9	Fast Fuzzy clustering Fast Fuzzy clustering based on kernel	85.26% 89.74%	2019	[10]

It therefore requires additional research to build advanced level models which include deep learning & machine learning algorithms that can provide the best accuracy of predictions of breast cancer and can be generalized for any type of dataset being used.

#### Application of CNN Model to predict Breast Cancer

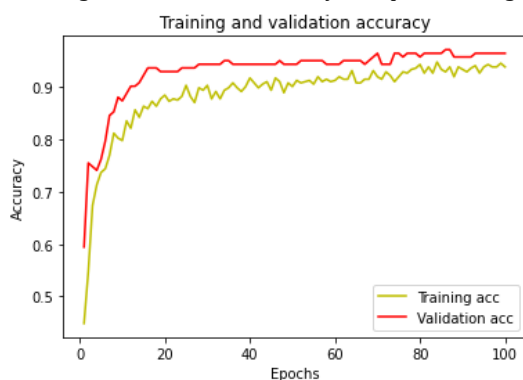
We used the Breast Cancer Wisconsin (Diagnostic) dataset from the UCI Machine Learning Repository [21], which had features (attributes) and corresponding labels (diagnosis) as a csv file. A digitized image of a fine needle aspirate (FNA) of a breast mass is used for computing the features. Given it is 2 class problem (B – benign or M – malignant), we built a CNN model to predict Breast Cancer

accuracy which based on the literature research and comparative analysis was the supposedly the most optimum one. The dataset consisted of 569 instances and 32 attributes. As a part of data pre-processing, we first analyzed if there were any null or missing values. We then performed an 'Exploratory' data analysis to comprehend the data distribution. The distribution of malignant and benign instances is shown in figure 7.



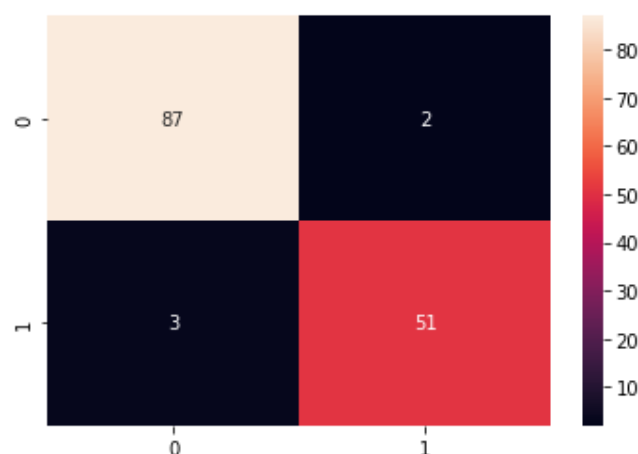
**Fig. 7: Data distribution of Malignant and Benign**

The dataset was then normalized and scaled to ensure the values brought to a similar range. The dataset was then split into training (75% of the data) and test (25% of the data). After having defined the training and test dataset, we built a CNN model by using Sequential as the model type. The input layer consists of 30 features (from the dataset), we use one hidden layer (dense layer) with 16 units) and an output layer. We randomly then drop 20% of connections between input and dense layer to generalize the model and not over fitting it. The single output in the output layer defines whether the cancer is malignant (value of 1) or benign (value of 0). We have used ReLu (rectified linear activation) as an activation function for input and dense layer and then applied 'sigmoid' activation to the output layer. The model is then compiled using the 'binary\_crossentropy' loss function and 'adam' optimizer which adjusts the learning rate throughout training. To make things even easier to interpret, we have used the 'accuracy' metric to see the accuracy score on the validation set when we train the model. We have then trained the model by cycling through the data for 100 times (epochs). We achieved an accuracy of 96.5% on the validation dataset. We finally used the model to predict the accuracy on the test dataset by plotting the confusion matrix. Out of the 54 malignant cases in the test dataset, the model was able to predict 51 of them correctly and out the 89 benign cases, it was able to predict 87 of them accurately. Training and validation accuracy is depicted in Figure 8.



**Fig. 8: Training & Validation accuracy post 100 epochs**

The model's accuracy on the test dataset for malignant cases was 94.4% and for benign cases was 97.8%. Confusion matrix for breast cancer prediction is presented in Figure 9.



**Fig. 9: Confusion Matrix plot for Breast Cancer prediction**

## Conclusion

In this paper, we have reviewed different algorithms associated with machine learning and deep learning that can be used for the prediction of breast cancer. Our main intent is to predict the occurrences of breast cancer more effectively and with improved accuracy using most optimal and suitable algorithm. This paper is also an effort to highlight all previous studies of machine learning algorithms that are being used for prediction of breast cancer. This paper provides all required and necessary information to the researchers who are working to establish the most optimum DL/ML algorithms for breast cancer detection and prediction. The review of this paper includes understanding of the various types of breast cancer, their symptoms, causes, current detection and diagnosis methods. The paper then covers the concepts of machine learning and deep learning specially focusing on techniques used for breast cancer prediction and detection. The paper discusses the CNN architecture in detail and outlines the fundamentals of developing an effective & efficient ML/DL model. Based on the past work performed between 2016 - 2020, the paper provides a comparative summary of ML techniques for breast cancer prediction. We used the Sequential DL model (classifier) on the Breast Cancer Wisconsin (Diagnostic) dataset and it yielded an accuracy of 96.5% on the validation dataset. The model had an accuracy of 94.4% to predict malignant cases on the test dataset and 97.8% to predict benign cases. In the future work, it is expected that certain data augmentation techniques may be used by the researchers to address issues of limited dataset availability. Researchers should also consider the issue of inequality of positive and negative data, as it can lead to biasedness towards positive or negative prediction. Another important issue that needed to be potentially addressed is the imbalanced number of breast cancer images against affected patches for correct diagnosis and prediction of breast cancer.

## Funding

None

## Acknowledgement

None

## Conflict of Interest

No Conflict of Interest

## Informed Consent

All authors have provided their consent

## Ethical Statement

Not required

## Author Contribution

All authors have equally contributed

## References

- Hazra, A., Mandal, S.K., Gupta, A.: Study and analysis of breast cancer cell detection using Naïve Bayes, SVM and Ensemble Algorithms. *Int. J. Comput. Appl.* 145, 0975–8887 (2016)
- Pritom, A.I., Munshi, M.A.R., Sabab, S.A., Shihab, S.: Predicting breast cancer recurrence using effective classification and feature selection technique. In: 19th International Conference on Computer and Information Technology (ICCIT), pp. 310–314. IEEE (2016)
- Asri, H., Mousannif, H., Al, M.H., Noel, T.: Using machine learning algorithms for breast cancer risk prediction and diagnosis. *Procedia Comput. Sci.* 83, 1064–1069 (2016)
- Mohammed, S.A., Darrab, S., Noaman, S.A., Saake, G. (2020). Analysis of Breast Cancer Detection Using Different Machine Learning Techniques. In: Tan, Y., Shi, Y., Tuba, M. (eds) *Data Mining and Big Data. DMBD 2020. Communications in Computer and Information Science*, vol 1234. Springer, Singapore. [https://doi.org/10.1007/978-981-15-7205-0\\_10](https://doi.org/10.1007/978-981-15-7205-0_10)
- Silva, J., Lezama, O.B.P., Varela, N., Borrero, L.A. (2019). Integration of Data Mining Classification Techniques and Ensemble Learning for Predicting the Type of Breast Cancer Recurrence. In: Miani, R., Camargos, L., Zarpelão, B., Rosas, E., Pasquini, R. (eds) *Green, Pervasive, and Cloud Computing. GPC 2019. Lecture Notes in Computer Science()*, vol 11484. Springer, Cham. [https://doi.org/10.1007/978-3-030-19223-5\\_2](https://doi.org/10.1007/978-3-030-19223-5_2)
- Anuranjeeta, K. K. Shukla, A. Tiwari, and S. Sharma, "Classification of Histopathological Images of Breast Cancerous and Non Cancerous Cells based on Morphological Features," *Biomedical and Pharmacology Journal*, vol. 10, no. 1, pp. 353–366, 2017.
- A.-A. Nahid, M. A. Mehrabi, and Y. Kong, "Histopathological Breast Cancer Image Classification by Deep Neural Network Techniques Guided by Local Clustering," *BioMed Research International*, vol. 2018, pp. 1–20, 2018.
- Alzubaidi, O. Al-Shamma, M. A. Fadhel, L. Farhan, J. Zhang, and Y. Duan, "Optimizing the Performance of Breast Cancer Classification by Employing the Same Domain Transfer Learning from Hybrid Deep Convolutional Neural Network Model," *Electronics*, vol. 9, no. 3, p. 445, 2020.
- Ray, A. A. Abdullah, D. K. Mallick, and S. Ranjan Dash, "Classification of Benign and Malignant Breast Cancer using Supervised Machine Learning Algorithms Based on Image and Numeric Datasets," *Journal of Physics: Conference Series*, vol. 1372, p. 012062, 2019.
- Classification of Breast Cancer using Fast Fuzzy Clustering based on Kernel," *IOP Conference Series: Materials Science and Engineering*
- IBM Cloud Education, "What is Machine Learning?," *Ibm.com*, 2015. [Online]. Available: <https://www.ibm.com/cloud/learn/machine-learning#toc-real-world-Lyja9GSr>.
- "Supervised learning and other machine learning tasks", *SuperAnnotate Blog*, 2021. [Online]. Available: <https://blog.superannotate.com/supervised-learning-and-other-machine-learning-tasks/>.
- J. Delua, "Supervised vs. Unsupervised Learning: What's the Difference?," *Ibm.com*, 2021. [Online]. Available: <https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning>.
- "AI & Machine Learning", *Simplilearn*, 2022. [Online]. Available: <https://www.simplilearn.com/tutorials/deep-learning-tutorial/deep-learning-algorithm>.
- Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, 2021.
- Ferlay, M. Colombet, I. Soerjomataram, C. Mathers, D. M. Parkin, M. Piñeros, A. Znaor, and F. Bray, "Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods," *International Journal of Cancer*, vol. 144, no. 8, pp. 1941–1953, 2018.
- "Statistics of Breast Cancer In India | Cytecare Hospitals", *Cytecare Hospital in Bangalore*, 2022. [Online]. Available: [https://cytecare.com/blog/statistics-of-breast-cancer/#:~:text=One%20in%20twenty%2Deight%20Indian,group%20\(1%20in%2060\)](https://cytecare.com/blog/statistics-of-breast-cancer/#:~:text=One%20in%20twenty%2Deight%20Indian,group%20(1%20in%2060)).
- "Understanding Your Breast Cancer Diagnosis", *Cancer.org*. [Online]. Available: <https://www.cancer.org/cancer/breast-cancer/understanding-a-breast-cancer-diagnosis.html>.
- "Breast Cancer Early Detection and Diagnosis | How To Detect Breast Cancer", *Cancer.org*. [Online]. Available: <https://www.cancer.org/cancer/breast-cancer/screening-tests-and-early-detection.html>.
- "Breast Cancer Overview: Causes, Symptoms, Signs, Stages & Types", *Cleveland Clinic*. [Online]. Available: <https://my.clevelandclinic.org/health/diseases/3986-breast-cancer>.
- UCI Machine Learning Repository (Center for Machine Learning and Intelligent Systems) [https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Diagnostic\)](https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic))
- A. Doran, *Cambridgeinsights.com*, 2020. [Online]. Available: <https://www.cambridgeinsights.com/post/data-science-in-asset-management-a-game-changer>.