A Review On Breast Cancer Using Machine Learning

Author:JAYA MEHRA, Department of Computer Science & Engg. , UIT RGPV BHOPAL (M.P.)
Dr. MANISH KUMAR AHIRWAR, Associate Professor, Department of Computer Science & Engg. , UIT RGPV BHOPAL (M.P.),
Dr. RAJEEV PANDEY, Associate Professor, Department of Computer Science & Engg. , UIT RGPV BHOPAL (M.P.),India,

Abstract—Breast cancer is the only type of cancer that affects women worldwide, and it may be a common cause of death. This paper's main goal is to develop a model for predicting breast cancer using several machine learning techniques, classification algorithms like Decision Tree (DT), Random Forest (RF), Logistic Regression (LR), Naive Bayes Gaussian (NB), On the other side, The purpose of this research is to estimate the likelihood that a patient will experience a recurrence of breast cancer. To improve the predictive performance of the Random Forest and Deep Neural Network classifiers, the researchers used them separately. Decision Tree(CART),Support Vector Machine (SVM), and Naïve Bayes for numerical datasets whose features are obtained from digitized images of breast mass, this paper study aims to improve accuracy in cancer database analysis and forecasting. The dataset is going to be examined, Since the UCI Machine Learning Repository makes the breast cancer dataset publicly available, the implementation phase dataset will be divided into 80% for the training phase and 20% for the testing phase. Utilize the algorithms for machine learning, assessed, and the model is trained during the procedure. Finally, prediction will be done using both numerical and picture test data.

Keywords – Breast Cancer, Machine Learning, Classification, Accuracy Precision, Random Forest, Decision Tree, Logistic regression

1. INTRODUCTION

One of the worst illnesses/diseases in India at the moment is breast cancer, which claims a lot of lives. Women are experiencing an increasing number of cancer cases every day as a result of dietary and lifestyle changes. It ranks as the second most frequent cause of death worldwide for women [1], Based on the collected data, this applies the ideas of machine learning (ML) and deep learning (DL) to predict breast cancer. The aberrant growth of fatty and fibrous tissues causes this malignancy, and the dissemination of cancer cells throughout the tissue is what causes the various stages of the disease [4].This is one of the most prevalent cancers that strike women, however compared to breast cancer, other malignancies and their patients can benefit substantially from treatment.. For picture datasets, this learning strategy is acknowledged as the most effective way to predict and categorize.

Even though earlier data classification techniques were less accurate, they were nonetheless employed since they allowed for accurate categorization and prediction. Features and hidden features are extracted using machine learning approaches based on numerical datasets and deep learning algorithms. The stride function, which is used to extract features from images of various sizes, yields the convolution value. CNN produces accurate results for the dataset that we used in this [13-17].While new machine learning approaches and algorithms are based on the building of models, this conventional regression-based method identifies the presence of of cancer using regression, despite the availability of new algorithms.

The model is intended to anticipate unknown data and provide a good expected outcome during its training and testing phases [19].The three main approaches in machine learning are preprocessing, feature selection or extraction, and classification [20].For the purpose of predicting and diagnosing cancer, the machine learning method's feature extraction component is essential. Regression is the basis for the suggested method's indication of cancer presence until better algorithms become available. A model that is intended to forecast new data and should perform well during the training and testing stages [19]. Here there are 3 main steps preprocessing features, extraction and classification. Figure 1, 2, shows the types of breast cancers, in this paper we consider IDC. Mammography mass lesion labeling is a common source of difficulty for radiology practitioners, which can result in unnecessary and expensive breast biopsies. The outcomes demonstrated that PCNN outperforms other methods for the same publically available datasets when combined with CNN. According to federal law, patients must be informed if the breast tissue on a mammogram is dense.
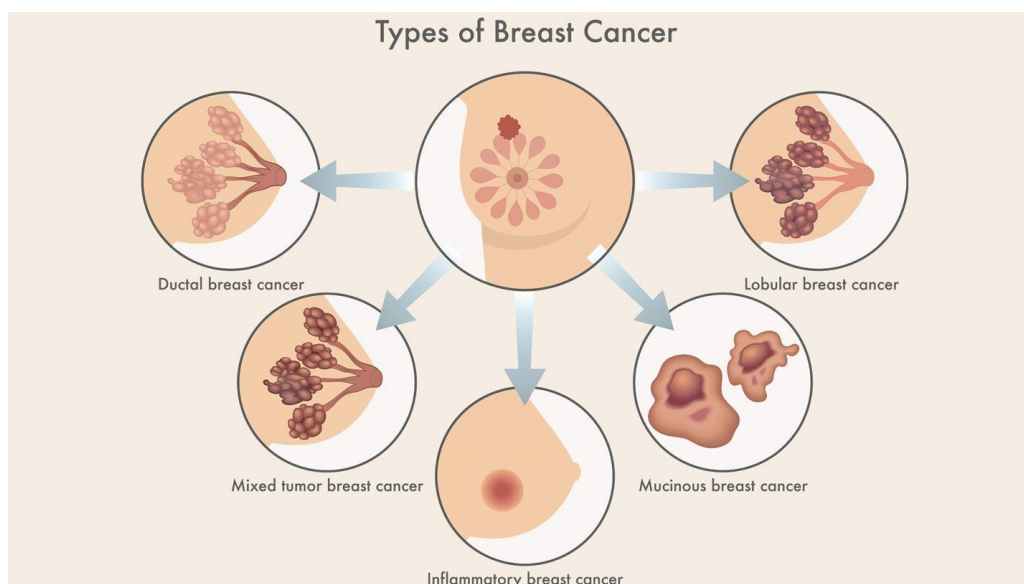
Figure.1 The various kinds of breast Cancer

Intelligent human services frameworks make extensive use of machine learning methods.especially for estimating and diagnosing breast cancer. While there are many machine learning classification and algorithms available for predicting the course of breast cancer, in this study we compare several types of classification techniques, including Support Vector Machine, Logistic Regression, k Nearest Neighbors, and Gaussian Naïve Bayes. Additionally, evaluate and contrast the various classifiers' performances in terms of accuracy, precision, recall, f1-Score, and Jaccard index. The results of this study offer an overview of the state of contemporary machine learning techniques for breast cancer screening.

The goal of this research is to develop a deep learning-based model that is effective. With the use of cutting-edge models, the suggested model may identify breast cancer in computerized mammograms with different densities and compare the outcomes.. For breast cancer, we were able to acquire varying specificity and good sensitivity. The diagnosis was made via dynamic contrast-enhanced (DCE) magnetic resonance imaging. This offers details on the morphology and functionality of the lesion.

FIGURE 2:
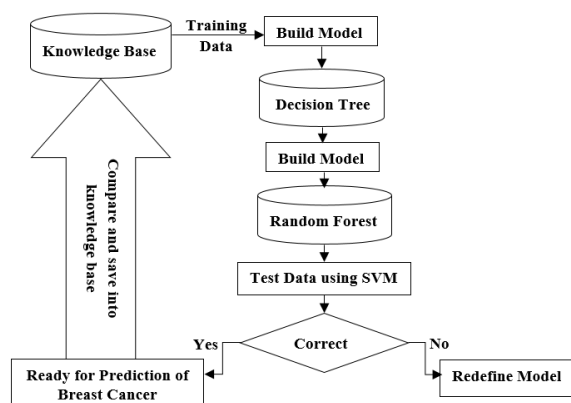
Numerous imaging techniques are available to detect and treat breast cancer in its early stages and reduce the number of deaths associated with the disease. Many assisted breast cancer detection techniques are also designed to improve the accuracy of symptoms Table 1 provides a succinct summary of several diagnostic methods with their advantages and disadvantages. This provides readers with a quick reference to understand the current status of breast cancer diagnosis and the need for an improved diagnostic approach.

1.1. Reasoning

Machine learning-based breast cancer diagnosis has been driven by the expectation that it could improve patient outcomes, decrease the disease's global impact, and assist in the creation of state-of-the-art medical research and technologies.

1.2. Advantages of This Study

The research has the following advantages: it enhances personalized therapy and early diagnosis. Furthermore, it can transform the way breast cancer is managed and save lives by impacting areas like research, affordability, and Access to healthcare services is available everywhere. The capabilities of machine learning in breast cancer diagnosis allow for lower healthcare expenses and a more positive impact on the management of breast cancer globally.

This is how the remainder of the paper is structured. We go over the earlier studies in the related work area in area 2. We introduced our suggested methodology in Section 3. We go over the accomplished experimental results in Section 4. The study's result is presented in Section 5

Table 1. Existing diagnostic method, advantages, and limitations

Existing Diagnostic Method	Advantages	Limitations
Mammography	Well-established	Limited sensitivity in dense breast tissue
	Widely accessible	False positives/negatives
	Detects structural changes and calcifications	False positives/negatives
Ultrasound	No radiation	Limited specificity
	Useful for dense breasts	Operator-dependent
	Differentiates cysts from solid masses	Limited detection in deep tissues
MRI (Magnetic Resonance Imaging)	High sensitivity	High cost
	No radiation	Longer exam duration
	Detailed soft tissue visualization	Requires specialized expertise to detect benign lesions
Biopsy (Fine Needle Aspiration or Core Needle Biopsy)	Provides tissue samples for definitive diagnosis	Invasive and uncomfortable
	High diagnostic accuracy	Small risk of complications
		Requires skilled medical staff
		Sample may not be representative
Clinical Breast Examination (CBE)	No radiation	Limited sensitivity
	Low cost	Dependent on examiner's expertise
	Can detect palpable masses	May miss non-palpable masses
Genetic Testing (BRCA1/BRCA2 Testing)	Identifies genetic mutations linked to increased risk	Applicable to specific subsets of patients
	Enables targeted prevention and treatment strategies	Limited to hereditary breast cancer cases

2. Literature review:

ML (machine learning) approaches have been employed in a number of studies in the healthcare arena in recent years to identify BC. Other scientists have applied the algorithms to difficult problems because they produce satisfactory results [12]. With an accuracy of almost 88%, a CNN algorithm was used to detect and diagnose invasive ductal carcinoma in breast cancer pictures [13, 14]. Furthermore, it is frequently employed in the medical industry to foresee and diagnose unusual events in order to gain a better knowledge of incurable illnesses like cancer [15]. Imaging and genetics-based breast cancer screening techniques have been the subject of numerous investigations.

Using deep learning techniques like convolutional neural networks (CNNs) can greatly improve the detection and diagnosis of breast cancer at an early stage, leading to better outcomes for patients. In one study, researchers examined different algorithms to predict non-communicable diseases (NCDs), like diabetes and cancer. They found that despite noisy and irrelevant data in NCD datasets, algorithms like KNN, SVM, and NN (neural networks) showed impressive resilience. These algorithms were tested on various datasets and were effective in identifying different health disorders.[20]

Another study looked at natural inspiration computing (NIC) approaches, inspired by insects, to diagnose diseases like breast, lung, prostate, and ovarian tumors. They developed algorithms that successfully recognized these tumors and improved breast cancer diagnosis by combining directed ABC (artificial bee colony) with neural networks. Additionally, they identified effective techniques for identifying diabetes and leukemia. Integrating NICs with conventional classification techniques produced more reliable results, but further research is needed, especially for detecting diseases at different stages.[21]

In another study, researchers found that neural networks, particularly when used in the early stages of illness, can classify cancer diagnoses accurately. However, preprocessing images for this method requires significant computing power. Overall, utilizing CNNs and artificial intelligence can help overcome these challenges in disease diagnosis.

By strengthening low-contrast features, lowering noise, eliminating artifacts, and maximizing image registration, CNNs and AI can raise the quality of medical images. Additionally, they can help with picture, segmentation, and ROI identification, allowing for accurate diagnosis and analysis of lesions or anatomical structures. AI systems can modify the brightness and contrast of images. AI algorithms help align images perfectly, while segmentation and ROI (region of interest) detection allow for precise examination of specific areas. Additionally, CNNs can upscale images through super-resolution techniques, using deep learning to generate sharper, more detailed images from lower-quality inputs, giving better diagnostic information.[21]

Computational intelligence approaches like fuzzy systems, artificial neural networks, and swarm intelligence, as well as evolutionary computing techniques like genetic algorithms, classifiers, and support vector machines, are useful in the field of smart health (Al-Antari, Al-Masni et al., 2018). [22]. According to studies in Khan, Khan et al., 2020) [21], the proposed CNN Improvements for Breast Cancer Classification (CNNI-BCC) model aids medical professionals in identifying breast cancer. The recommended approach classifies breast cancer subtypes using a trained deep learning neural network system. Based on data from 221 real patients, the results have a 90.50 percent accuracy rate.

This model is able to classify and identify lesions related to breast cancer without the requirement for human interaction .This model's evaluation demonstrates that it can assess affected patients' circumstances during the detection phase, demonstrating that it is superior to .the condition of affected patients during the detecting stage, demonstrating that it is an advancement over previous methods (Tanabe, Ikea et al., 2020). [23]

By applying Naive Bayes, Support Vector Machine, Neural Network, and Decision Tree approaches to a Wisconsin breast cancer development dataset, Chaurasia et al. [7] examined the effectiveness of supervised learning classifiers. 96.84% was the precise conclusion that the Support Vector Machine provided to the chief based on the investigation results, which were deemed reliable. On a mass, it can be "circumscribed," "micro-lobulated," "obscured," "indistinct," or "speculated," as illustrated in Figure 2. Nevertheless, early identification of breast cancer is essential to reducing the number of cancer deaths and improving patients' quality of life.

TABLE 2 .
A LITERATURE REVIEW FOR BREAST CANCER PREDICTION USING MACHINE LEARNING

ALGORITHM	TOOL	DATA SET	DATA PREPROCESSING MEHTOD	EVALUTION	VALIDATION TECHNIQUE	ACCURACY
SVM KNN LG	MATLAB	UCI DEPOSITORY WDBC WPBC	FEATURE SELECTION	FUCTION GRAPH(LR)DISTRIBUTION KERNAL DISTRIBUTION(NB)	SVM DYNAMIC KNN EUCLIDEAN LR GENERATION NB KERNEL	93% 95% 90% 92%
SVM KNN LR	SPYDER	UCI DEPOSITIRY	FEATURE SECLETION DIMENSIONALLY REDUCTION	PRINCIPLE COMPONENT ANALYSIS(PCA)	SVM-PCA TESTING KNN EUCLIDEAN LR LOGISTIC FUCTION	92.78% 92.23% 92.10%
DT KNN LR	WEKA	IRANIAN CENTRE ICBC	DATACLEANING AND PREPARATION	EXPECTATION MAXIMIZATION ALGORITHM(EMA)	10 FOLD CROSS VALIDATION	93.6% 94.7% 95.7%
MLP KNN CART NB SVM	MATHEMATCAL	UCI DEPOSITORY	STANDARDIZE RESCALING METHOD	BINARY CLASSIFICATION ACCURACY METHOD	10 FOLD CROSS VALIDATION	99.12% 95.6% 93.85% 94% 98%
SVM CART NB KNN	WEKA	WBC WISCONSIN BREAST CANCER	KAPPA STATISTIC MEAN ABSOLUTE ERROR	CONFUSION MATRIX	10 FOLD CROSS VALIDATION	95% 97% 95% 95.27%
MB ANN CART	WEKA	SEER DATABASE	FEATURE FUNCTION	CONFUSION MATRIX	10 FOLD CROSS VALIDATION	84% 86% 86.7%
R F ALGORITHM CART TREE BASE	WEKA	ANTEMA DATABASE	FEATURE EXTRACTION	BINARY CLASSIFIFCATION ACCURACY METHOD	10 FOLD CROSS VALIDATION	92.2% 90% 90.9% 90% 90%
J48 DT ZERO R	WEKA	PATHOLOGY REPORT	KAPPA STATISTIC ABSOLUTE ERROR	CONFUSION MATRIX	CROSS VALIDATION	95.9% 1%
J48 ALGORITHM ADT(ALTERING DECISION TREE)	WEKA	INDIA CANCER INSTITUTE ADYAR CHANNAI	FEATURE SELECTION	BINARY CLASSIFICATION ACCURACY METHOD	WEIGHTED AVERAGE PF PARAMETRES	98.10% 97.70% 98.50%

NB J48 DT	WEKA	LASUTH CANCER DATASET NIGERIA	FEATURE EXTRACTION	CONFUSION MATRIX	PERFORMANCE EVALUTION MODEL	82.6% 9.2%

FIGURE 3. Edges of breast mass [6]

Circumscribed   Obscured   Microlobulated   Ill-defined   Spiculated
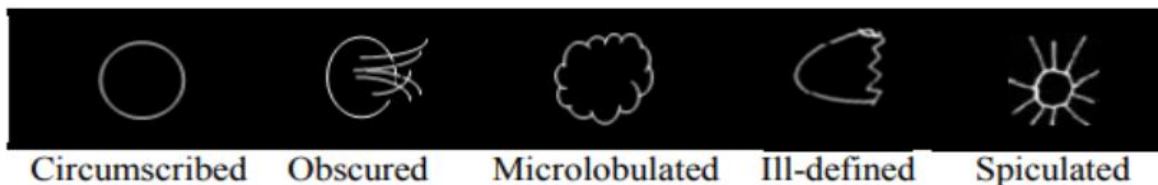
Age, heredity, female sex, and having thick breasts are the most common and well-known risk factors for breast cancer. Breast density is a metric used to represent the size of the unique tissues that make up a woman's breasts as well as the appearance of the breasts on a mammography. Women with dense breasts are more likely to get breast cancer, and there is a correlation between breast density and age, with younger women having denser breasts than older women. [3]

Breast Imaging Data and Reporting System, or BI-RADS, was developed by the American College of Radiology [9]. BI-RADS evaluation categories are shown in Figure 3. It is helpful for radiologists to be encouraged by "BI-RADS" to think about which category is most appropriate.
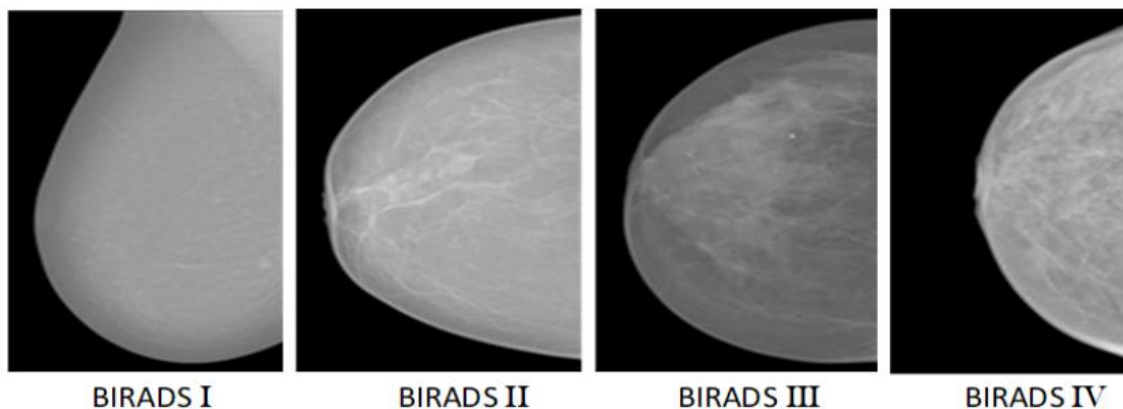

BIRADS I   BIRADS II   BIRADS III   BIRADS IV

FIGURE 4.

3. METHODOLOGY :

3.1 DATA SET DISCRIPTION:

The numerical dataset was provided from the WBCD repository. Features consist of a breast mass fine needle aspirate (FNA). To characterize the features of the cell nuclei seen in the scanned pictures, 30 features were retrieved. Of the 569 patients in the sample, 212 had a malignant outcome, and 357 had a benign outcome. The dataset's classes are divided into two or four groups, where two represent the benign case and four represent the malignant case.

The dataset used in this paper is freely available from the UCI Machine Learning Repository [2]. The collection consists of several hundred human cell test records, each containing an estimation of a group of cell

characteristics. The dataset with the following qualities as a result:.

i Clump of ID Number Depth
ii. Consistency in Cell Size
iii. Consistency in Cell Form
iv. Adhesion at the Margin
v. Size of a Single Epithelial Cell
vi. Unadorned Nuclei
vii. Chromatin Bland
viii. Typical Nucleoli
ix. Mitoses
x. Class

The patient identifiers are contained in the ID Number attribute. Each patient's unique cell test characteristics are included in the attribute Clump Thickness to Mitoses. The values are ranked from 1 to 10, where 1 is the value that is closest to the beginning. The

class field gives the conclusion on whether the tests are benign (value = 4) or malignant (value = 2), as confirmed by isolated clinical procedures.

Table 3. The dataset's class statistics are displayed

Class	Instances	% Distribution
Begin	458	65.52
Malignant	241	34.48
Total	699	100

3.2 DATA PRE- PROCESSING:
"Data pre-processing" is the phrase used to describe the steps involved in transforming unstructured data into structured data, resizing, and eliminating unwanted material from a dataset. The mean value is used to fill in the missing traits in the dataset. To make sure the data is distributed correctly, a random selection of the dataset is made.[27]

3.3 PHASE OF TRANING AND TESTING:
The testing step will provide fresh data to be analyzed in order to assess how well our algorithm performs and behaves in terms of prediction, while this phase extracts the features from the dataset. As mentioned earlier, there are two parts to the dataset. The process of cross-validation is used to prevent fitting. Every version of our method portions data using a ten-fold method, with the remaining nine fold being used for testing and training.[28]

4. RESUL AND DISCUSSION:
Using the evaluation measures EDA dataset, we have presented a standardized approach in this study for comparing the state-of-the-art. Initially, we gathered the information from kaggle.com (accessed on April 15, 2023). For the simulation, we utilized Python Jupiter Notebook 6.4.12. The initial findings of the studies of explanatory data using various models are shown in Table 3. Python extensions like Pandas, Seaborn, Plotly, and Bokeh are examples of EDA resources. Pandas is a feature-rich data analysis package that provides Data Frames and Series. Plotly is a versatile framework for making interactive visualizations including dashboards and 3D plots, whereas Seaborn is a high-level interface for making statistical visualizations adaptable framework for building 3D plots and dashboards, among other interactive visualizations. A library called Bokeh is used to create online interactive dashboards. The predictions are applied to the dataset's subsets in this graph. Thirty percent is the test data and seventy percent is the training data for each set. The table included comparisons with radius_mean, id, and other metrics.

Table 4.Exploratory data analysis

(a)

id	Diagnosis	Radius_Mean	Texture_Mean	Perimeter_Mean	Area_Mean	Smoothness_Mean	Compactness_Mean	Concavity_Mean	Concave Points_Mean	Radius_Worst
0	842302 M	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.3001	0.14710	25.38
1	842517 M	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.0869	0.07017	24.99
2	84300903 M	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.1974	0.12790	23.57
3	84348301 M	11.42	20.38	77.58	386.1	0.14250	0.28390	0.2414	0.10520	14.91
4	84358402 M	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.1980	0.10430	22.54

(b)

Texture_Worst	Perimeter_Worst	Area_Worst	Smoothness_Worst	Compactness_Worst	Concavity_Worst	Concave Points_Worst	Symmetry_Worst	Fractal_Dimension_Worst	Unnamed: 32
17.33	184.60	2019.0	0.1622	0.6656	0.7119	0.2654	0.4601	0.11890	NaN
23.41	158.80	1956.0	0.1238	0.1866	0.2416	0.1860	0.2750	0.08902	NaN
25.53	152.50	1709.0	0.1444	0.4245	0.4504	0.2430	0.3613	0.08758	NaN
26.50	98.87	567.7	0.2098	0.8663	0.6869	0.2575	0.6638	0.17300	NaN
16.67	152.20	1575.0	0.1374	0.2050	0.4000	0.1625	0.2364	0.07678	NaN

Because of extensive screening systems and public education campaigns, the percentage is frequently higher in wealthy countries than in developing ones because of limited resources and lower awareness. The American Cancer Society (ACS), the National Cancer Institute (NCI), the World Health Organization (WHO), and other reputable organizations are great places to find reliable and current information about breast cancer.
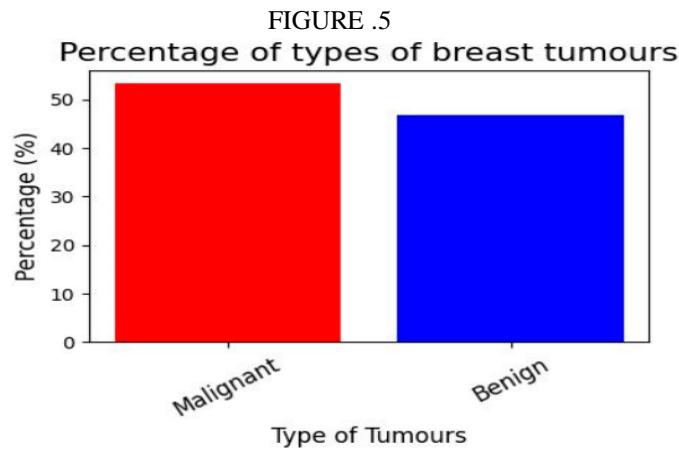
FIGURE .5

The diagnosis, radius_mean, testure_mean, perimeterr_mean, area_mean, smoothness_mean, etc. are among the variables on the y-axis. No link is shown by a correlation coefficient between −1 and 1.at all, but a figure that is either positive or negative denotes a robust correlation between the two variables. It is crucial to remember that correlation can only quantify the linear relationship between variables. There is a correlation of at least 69% between each of these variables and the patient's prognosis. Diagnostics 2023, 13, 3113 16 of 22 that have a correlation-based measurement. There is a correlation of at least 69% between each of these variable and the prospects for the patient.
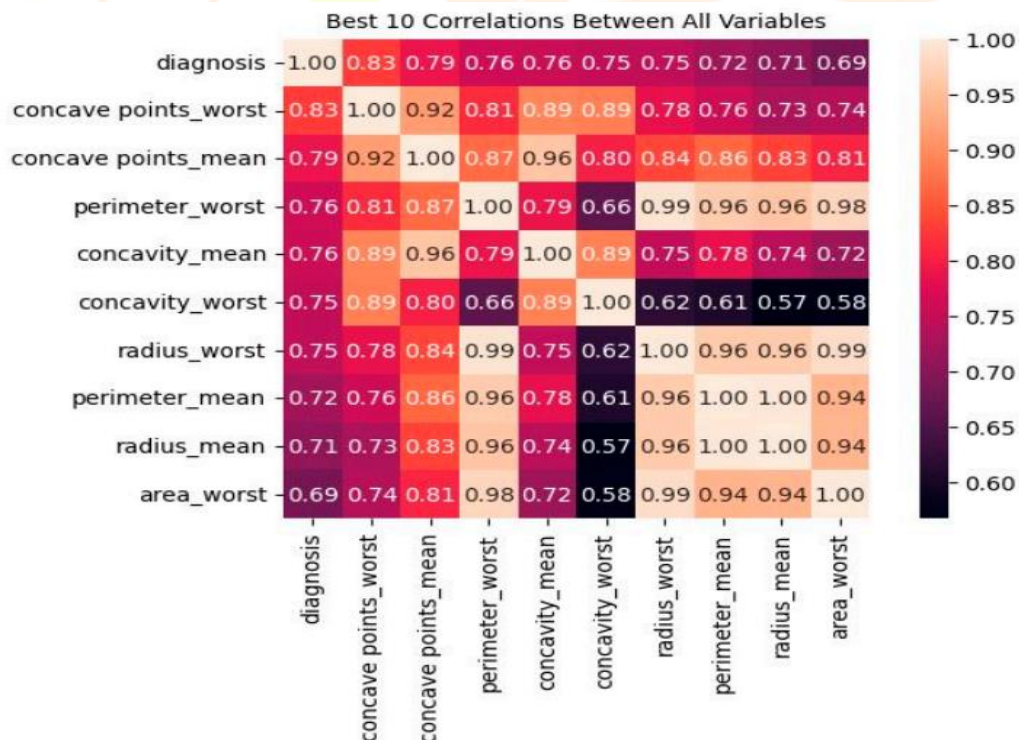

FIGURE 6 .Correlation among the top 10 variables.

ML has the potential to significantly enhance breast cancer detection and diagnosis. Nonetheless, there are legal, ethical, and technological barriers to be addressed before technology may be effectively incorporated into clinical practice. As science and technology grow, the goal is to create a more accurate, user-friendly, and patient-centered approach to breast cancer detection and management.

	Model	Scores
0	Random Forest Classifier	96.491228
3	Decision Tree	93.859649
1	Logistic Regression	92.982456
2	KNeighbour Clasifier	92.105263
5	Linear SVC	89.473684
4	SVC	87.719298

FIGURE 7 CLASSIFIER ACCURACY RESULT

5. CONCLUSION AND FUTURE WORK:

In this work, we have compared the four Machine Learning classification parameters. algorithms on the UCI Machine Learning Repository Wisconsin Wisconsin breast cancer dataset, especially k Nearest Neighbor, Support Vector Machine, Logistic Regression, and Naive Bayes. Finding the most highly accurate machine learning algorithm to use as a diagnostic tool for breast cancer was the goal of this comparative investigation. For the given dataset, k Nearest Neighbor has the highest accuracy, which is in line with the forecast findings. This demonstrates that when predicting breast cancer, k Nearest Neighbor consistently outperforms Support Vector Machine, Logistic Regression, and Naive Bayes.

This study's main goal is to develop and implement a novel algorithm for deciphering and arranging data on chest diseases from mammography and pathology findings from patient images that are retrieved from the cloud of UCI repositories. Convolutional neural association modeling in Python programming is used to do this, and the outcomes are verified. The results show that the proposed CNN performs better than guesses in identifying and requesting breast cancer for picture datasets. Additionally, SVM has demonstrated superior performance over CART, NB, and KNN in the analysis and numerical dataset prediction of cancer.

Artificial intelligence developments have improved the accuracy of mammography, and deep learning models are being created to identify breast cancer in digital mammograms. Convolution neural networks and artificial intelligence (AI) are being used in healthcare to enhance image processing and lessen the need for human eye identification. Breast MRI is a sensitive imaging method with exceptional sensitivity and specificity.

Future studies on the use of ML for breast cancer diagnosis may look into these and other options. In order to significantly advance the identification and treatment of breast cancer, more research and collaboration between data scientists, physicians, and Researchers are vital.

REFERENCE:

1.   Chen, H., Wang, N., Du, X., Mei, K., Zhou, Y., & Cai, G. (2023). Classification prediction of breast cancer based on machine learning. *Computational Intelligence and Neuroscience*, *2023*, 1–9. https://doi.org/10.1155/2023/6530719

2.   Ali, Y., Sardaraz, M., Tahir, M., Elmannai, H., Hamdi, M., & Ksibi, A. (2023). IDriveGenes: Cancer Driver Genes Prediction using Machine Learning. *IEEE Access*, *11*, 28439–28453. https://doi.org/10.1109/access.2023.3259907

3.   Lupat, R., Perera, R., Loi, S., & Li, J. (2023). MOANNA: Multi-Omics Autoencoder-Based Neural Network Algorithm for predicting breast cancer subtypes. *IEEE Access*, *11*, 10912–10924. https://doi.org/10.1109/access.2023.3240515

4.   Naseem, U., Rashid, J., Ali, L., Kim, J., Qazi, E., Awan, M. J., & Imran, M. (2022). An automatic detection of breast cancer diagnosis and prognosis based on machine learning using ensemble of classifiers. *IEEE Access*, *10*, 78242–78252. https://doi.org/10.1109/access.2022.3174599

5.   Ilyas, Q. M., & Ahmad, M. (2021). An Enhanced ensemble diagnosis of Cervical Cancer: A pursuit of machine intelligence towards Sustainable health. *IEEE Access*, *9*, 12374–12388. https://doi.org/10.1109/access.2021.3049165

6.   Ghorbani, R., Ghousi, R., Makui, A., & Atashi, A. (2020). A new hybrid predictive model to predict the early mortality risk in intensive care units on a highly imbalanced dataset. *IEEE Access*, *8*, 141066–141079. https://doi.org/10.1109/access.2020.3013320

7.   Fatima, N., Liu, L., Sha, H., & Ahmed, H. (2020). Prediction of breast cancer, comparative review of machine learning techniques, and their analysis. *IEEE Access*, *8*, 150360–150376. . https://doi.org/10.1109/access.2020.3016715

8.   Islam, M. M., Haque, M. R., Iqbal, H., Hasan, M., Hasan, M., & Kabir, M. N. (2020). Breast Cancer Prediction: A comparative study using machine learning techniques. *SN Computer Science*, *1*(5). https://doi.org/10.1007/s42979-020-00305-

9.   khuriwal, N., & Mishra, N. (2018). Breast cancer diagnosis using adaptive voting ensemble machine learning algorithm. *2018 IEEMA Engineer Infinite Conference (eTechNxT)*. https://doi.org/10.1109/etechnxt.2018.838535

10.   Asri, H., Mousannif, H., Moatassime, H. A., & Noël, T. (2016). Using machine learning algorithms for breast cancer risk prediction and diagnosis. *Procedia Computer Science*, *83*, 1064–1069. https://doi.org/10.1016/j.procs.2016.04.224

11.   Wang, Z., Yu, G., Kang, Y., Zhao, Y., & Qu, Q. (2014). Breast tumor detection in digital mammography based on extreme learning machine. *Neurocomputing*, *128*, 175–184. https://doi.org/10.1016/j.neucom.2013.05.053

12.   Wolberg, W. H., & Mangasarian, O. L. (1990). Multisurface method of pattern separation for medical diagnosis applied to breast cytology. *Proceedings of the National Academy of Sciences of the United States of America*, *87*(23), 9193–9196. https://doi.org/10.1073/pnas.87.23.9193

13.   Fatima, N., Liu, L., Sha, H., & Ahmed, H. (2020). Prediction of breast cancer, comparative review of machine learning techniques, and their analysis. *IEEE Access*, *8*, 150360–150376. https://doi.org/10.1109/access.2020.3016715

14.   Sengar, P. P., Gaikwad, M. J., & Nagdive, A. S. (2020). Comparative Study of Machine learning Algorithms for breast cancer prediction. *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*. https://doi.org/10.1109/icssit48917.2020.9214267

15.   Yao, D., Yang, J., & Zhan, X. (2013). A novel method for disease prediction: hybrid of random forest and multivariate adaptive regression SPlines. *Journal of Computers*, *8*(1). https://doi.org/10.4304/jcp.8.1.170-177

16.   National Academy of Sciences: Abstracts of papers presented at the annual Meeting, 24-26 April 1961, Washington, D.C. (1961). *Science*, *133*(3461), 1363–1370. https://doi.org/10.1126/science.133.3461.1363

17.   Lundin, M., Lundin, J., Burke, H. B., Toikkanen, S., Pylkkänen, L., & Joensuu, H. (1999). Artificial neural networks applied to survival prediction in breast cancer. *Oncology (Basel)*, *57*(4), 281–286. https://doi.org/10.1159/000012061

18.   Lg, A., & At, E. (2013). Using three machine learning techniques for predicting breast cancer recurrence. *Journal of Health & Medical Informatics*, *04*(02). https://doi.org/10.4172/2157-7420.1000124

19.   Ayer, T., Chhatwal, J., Alagöz, O., Kahn, C. H., Woods, R. W., & Burnside, E. S. (2010). Comparison of logistic regression and artificial neural network models in breast cancer risk estimation. *Radiographics*, *30*(1), 13–22. https://doi.org/10.1148/rg.301095057

20.   Tan, M., Zheng, B., Leader, J. K., & Gur, D. (2016). Association Between Changes in Mammographic Image Features and Risk for Near-Term Breast Cancer Development. *IEEE Transactions on Medical Imaging (Print)*, *35*(7), 1719–1728. https://doi.org/10.1109/tmi.2016.2527619

21.   Dhahri, H., Maghayreh, E. A., Mahmood, A., Elkilani, W. S., & Faisal, M. (2019). Automated breast cancer diagnosis based on machine learning algorithms. *Journal of Healthcare Engineering (Print)*, *2019*, 1–11. https://doi.org/10.1155/2019/4253641

22.   Κούρου, K., Exarchos, T. P., Exarchos, K., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal*, *13*, 8–17. https://doi.org/10.1016/j.csbj.2014.11.005

23.   Asri, H., Mousannif, H., Moatassime, H. A., & Noël, T. (2016). Using machine learning algorithms for breast cancer risk prediction and diagnosis. *Procedia Computer Science*, *83*, 1064–1069. https://doi.org/10.1016/j.procs.2016.04.224

24.   Geras, K. J., Mann, R. M., & Moy, L. (2019). Artificial intelligence for Mammography and Digital Breast tomosynthesis: Current concepts and future Perspectives. *Radiology*, *293*(2), 246–259. https://doi.org/10.1148/radiol.2019182627

25.   Sharma, S., Aggarwal, A., & Choudhury, T. (2018). Breast cancer detection using machine learning algorithms. *2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)*. https://doi.org/10.1109/ctems.2018.8769187

26.   Khuriwal, N., & Mishra, N. (2018). Breast cancer diagnosis using deep learning algorithm. *2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*. https://doi.org/10.1109/icacccn.2018.8748777

27.   Moons, K. G., Altman, D. G., Reitsma, J. B., Ioannidis, J. P. A., Macaskill, P., Steyerberg, E. W., Vickers, A. J., Ransohoff, D. F., & Collins, G. S. (2015). Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. *Annals of Internal Medicine*, *162*(1), W1–W73. https://doi.org/10.7326/m14-0698

28.   Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, *30*(7), 1145–1159. https://doi.org/10.1016/s0031-3203(96)00142-2

29.   Tseng, H. E., Wei, L., Cui, S., Luo, Y., Haken, R. T., & Naqa, I. E. (2018). Machine learning and imaging informatics in oncology. *Oncology*, *98*(6), 344–362. https://doi.org/10.1159/000493575

30.   Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M. A., Chou, K., Cui, C., Corrado, G. S., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine (Print)*, *25*(1), 24–29. https://doi.org/10.1038/s41591-018-0316-z