Volume 9, Issue 4, April – 2024 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR707

Unlocking the Potential Thorough Analysis of Machine Learning for Breast Cancer Diagnosis

[1] Dr. P. Bhaskar
Professor, Dept. of CSE
QIS College of Engineering and Technology(A), Ongole
Andhra Pradesh–523001, India

[2]Tahaseen Syed
Dept. of IT
QIS College of Engineering and Technology(A), Ongole
Andhra Pradesh–523001, India

[3] Hima Varsha Daka
Dept. of IT
QIS College of Engineering and Technology(A), Ongole
Andhra Pradesh–523001, India

[4] Nikhil Kumar Theegala
Dept. of IT
QIS College of Engineering and Technology(A), Ongole
Andhra Pradesh–523001, India

[5]Manikanta Tulluri
Dept. of IT
QIS College of Engineering and Technology(A), Ongole
Andhra Pradesh–523001, India

Abstract:- To forecast breast cancer with the goal of giving a thorough rundown of current developments in the area. Given that breast cancer is among the world's leading causes of mortality for women; improving patient outcomes requires early detection. This study looks into the ability to predict outcomes using a variety of machine learning (ML) models, including random forests, logistic regression, support vector machines, decision trees, k-nearest neighbours, and deep learning neural networks, in predicting the incidence of breast cancer from patient data, including genetic markers, imaging results, and demographics.

Aims to provide a comprehensive analysis of presetn advancements, obstacles, and prospects in the field of CNN-based techniques for breast cancer identification. The review begins by outlining the urgent need for reliable and accurate diagnostic methods for breast cancer, highlighting the critical role that early identification plays in enhancing patient outcomes. Which delves into the intricate architecture of CNNs, revealing its unique applicability to mammography image analysis as well as their innate advantages in image classification tasks. Important topics of discussion include the various CNN architectures used for two- and three-dimensional (2D) imaging methods used in breast cancer diagnosis.

Keywords:- *Breast, Cancer. Tumor, Resnet50 V2, VGG16, CNN.*

I. INTRODUCTION

Breast cancer is a serious worldwide hazards health concern that necessitates the development of the healthy lifestyle. Novel strategies for prompt detection and precise diagnosis. Convolutional neural networks in particular, have been increasingly prevalent in deep learning approaches, and their convergence has revolutionised breast cancer diagnosis in recent years. This thorough analysis attempts to offer a thorough examination of the uses, developments, difficulties, and prospects related to using CNNs in the work of breast cancer detection and diagnosis.

By utilising extensive patient data sets and advanced algorithms, machine learning (ML) models are capable of identifying different patterns and correlations insideh the data, leading to more precise and individualized risk assessments for breast cancer.

Artificial neural networks (ANNs), random forests, gradient boosting algorithms like XGBoost, and more sophisticated approaches like logistic regression and support vector machines are examples of machine learning (ML) algorithms that have been thoroughly investigated for their potential in analyzing various patient data sets to guess the likelihood of breast cancer occurrence.

Among the models now in use, logistic regression is popular for predicting tumour malignancy based on a variety of characteristics and provides a simple method for binary classification tasks. Support vector machines will be trained on characteristics obtained from medical imaging or patient data for breast cancer prediction. SVMs are well-known for their efficacy in categorizing data into two groups .. By combining several decision trees, an ensemble learning method known as random forests may forecast the

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR707

occurrence of breast cancer by accounting variables including tumor size, shape, and texture. Furthermore, the capacity of artificial neural networks—more specifically, deep learning models like CNN—to recognise complex patterns in medical data has drawn attention. This capability makes CNNs a viable option for image-based breast cancer screening.

Newer methods, including the effective gradient boosting algorithm XGBoost, have shown good performance and interpretability in a variety of medical applications, including the prediction of breast cancer, in addition to these well-established models. The suggested approach, which builds on these earlier models, uses convolutional neural networks (CNNs) to predict breast cancer. Specifically, it uses the VGG16 and ResNet50V2 architectures. These designs have demonstrated potential in precisely identifying malignant spots and are well-suited for feature extraction from medical pictures. The authors of work want to improve the accurate prediction of breast cancer using machine learning by assessing these CNN architectures' performance and contrasting it with more conventional ML algorithms.

These designs have demonstrated potential in precisely identifying malignant spots and are well-suited for feature extraction from medical pictures.

II. RELATED WORKS

With around 1 in 8 people ever being at danger. Although early discovery is essential for survival, it is still difficult to differentiate benign from malignant tumor's. To help with this endeavor, researchers are turning to machine learning. Eight machine learning techniques were used on the Breast Wisconsin Diagnostic dataset in a recent study. With an astounding accuracy of 98.08%, SVM and ANN were found to be the most successful.[1] This study emphasizes the value of early diagnosis in saving lives and the promise of machine learning in breast cancer detection.

The present approaches have limits in terms of resource needs and accessibility, but early identification is essential for increasing survival rates and lowering treatment costs. methods to classify image features of benign and aggressive breast cancer, including Random Forest, Support Vector Machine, and Artificial Neural Networks. Experiments conducted on the Wisconsin Breast Cancer dataset showed that the Artificial Neural Network was the most effective model, with a 99% classification accuracy rate.

Although early identification is essential for raising survival rates and cutting treatment costs, the available techniques have limitations in terms of accessibility and resource needs. In order to distinguish between benign and malignant breast cancer imaging characteristics, this study focuses on applying artificial intelligence approaches for early breast cancer detection.[3] Support Vector Machine, and ANN algorithms on the Wisconsin dataset of breast cancer, the study's classification accuracy was 99%.

The accuracy of detection of cancer has improved with advances in machine learning and technology.[4] Machine learning methods, such as Multi-layer Perceptron's, and Linear Regression, allow intelligent to learn from historical past data and identify patterns. The use of these approaches to Wisconsin data on breast cancer is surveyed in this study, and the most successful approach is the multilayer perceptron, which provides encouraging insights into cancer prognosis and improves detection.

Study investigates the use of two well-liked machine learning methods for the categorization of the Wisconsin Breast Cancer dataset. ROC Area values, accuracy, precision, recall, and overall method performance are compared. The strategy that performs the best and has the highest accuracy is Support Vector Machine.[5] This study demonstrates how machine learning might improve cancer detection and prognosis.

In order to diagnose breast cancer, machine learning approaches are essential.[6] In this research, an adaptive ensemble voting method is proposed. The study evaluates the efficacy of logistic algorithms and artificial neural networks (ANN) in combination with ensemble machine learning using the Wisconsin Breast Cancer database. The outcomes show that, in comparison to other machine learning algorithms, the ANN technique together with the logistic algorithm achieves an amazing 98.50% accuracy.

There are discrepancies in the diagnosis.[7] The high death rate associated with breast cancer has increased interest in computer-aided methods for early diagnosis and treatment. Artificial intelligence (AI) has the potential to support professionals in their decision-making by yielding precise outcomes. This research uses five breast cancer datasets for binary classification evaluation of four models. The results show that CNN is more accurate than other algorithms. Future research on breast cancer is guided by this work.

Breast cancer is responsible for about 2.3 million diagnoses and 0.7 million deaths per year, thus early diagnosis is critical. Unfortunately, manual diagnosis frequently fails to identify cancer in its early stages. [8]The study uses 30 unique features from patient records to classify tumors as malignant. To save lives through early detection, many machine learning techniques are used, with the classification model selected based on the greatest accuracy attained.

[9]improved prognosis and early detection of breast cancer. This study examines five machine learning algorithms: K-Nearest Neighbour, Decision Tree, Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), and SVM, using the Wisconsin Diagnostic Breast Cancer dataset. The goal is to identify the most effective algorithms for early breast cancer diagnosis.

Accurate diagnosis is crucial, yet methods like FNAC and mammography have interpretation uncertainty. An early diagnosis is necessary for therapy to be effective. Machine learning techniques for data mining and classification offer efficient result categorization. The Wisconsin Breast Cancer dataset is used in this study to evaluate the accuracy, AUC, and effectiveness of several machine learning strategies.[10] With an AUC of 99.4, the Support Vector Machine (SVM) model produced results with an accuracy of 96.25%. The findings of this study demonstrate how machine learning might improve breast cancer diagnosis and prognosis, offering promise for improved outcomes via early detection.

Theidentification of cancer using logistic regression, decision trees, random forests, and CNN is compared. A key component of this prediction process is machine learning. Three classification algorithms are used in the study, and their performance and accuracy are assessed. [11]It is essential to handle unbalanced class data and calls for appropriate preparation

Utilizing [12] the Wisconsin Diagnostic Data Set, this investigation illustrates fuzzy-based breast cancer diagnosis utilising machine learning techniques. When precision, accuracy, specificity, and recall are taken into account, the combined model performs better than the separate ML models. The efficacy of fuzzy-based SVM and DT in diagnosing breast cancer illnesses is demonstrated by their high accuracy of 98.2%, precision of 97.6%, recall of 96.5%, and specificity of 97.8%.

As a result of technological advancements, the medical community is investigating machine learning methods for automatically classifing the prediction breast cancer tumors as benign or malignant.[13] This study looks at many machine learning techniques for analysing and predicting the stages of breast cancer. The study intends to improve patient treatment and survival rates by emphasizing early detection and precise categorization, highlighting the need of prompt diagnosis in the fight against this deadly illness.

To increase patient survival rates, prompt identification is essential. The main diagnostic technique is mammography, however results may suffer from delays in diagnosis caused by the lack of specialists. By enhancing efficacy and efficiency in cancer diagnostics, machine learning provides a remedy.[14] In order to improve categorization and prediction accuracy, this study summarizes machine learning methods used in breast cancer detection. Ensuring accurate diagnosis and categorization is the goal in the battle against breast cancer, with the ultimate goal being improved patient outcomes.

Enhanced prognosis models for more efficient treatment arrangements. While early detection is important, further research is still required. The motive of this research is to form a long-lasting machine learning model for identifying the kind of breast cancer. [15] After analysing five machine learning algorithms, XGBoost showed the best results, with an impressive 95.42% accuracy along with great sensitivity, specificity, and F-1 score. Even though XGBoost seems promising, further study is required before it can be used commercially in the healthcare industry. This work paves the door for better patient outcomes by highlighting the significance of machine learning in improving breast cancer diagnosis and therapy.

III. DATASET DESCRIPTION

This dataset is made up of a number of photos that were analyzed using machine learning methods to identify breast cancer. There are 657 photos that are labelled as non-cancer and 587 images that are classified as cancerous. Every grayscale image has undergone preprocessing to guarantee consistency in terms of size and format. The dataset is separated into two primary groups.[16]

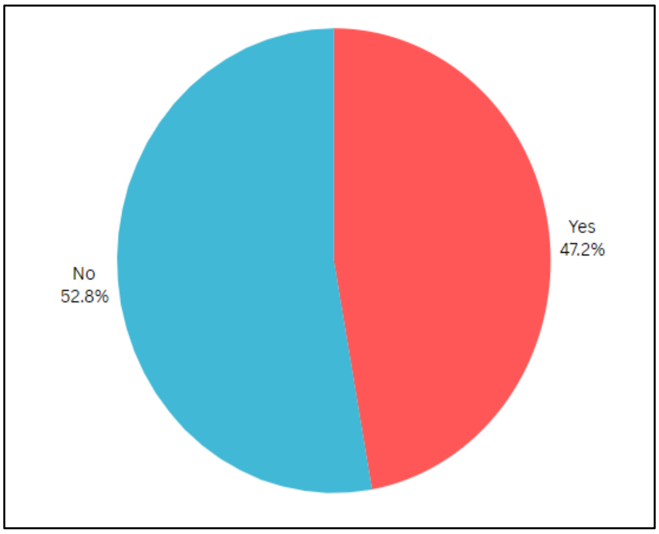

Fig. 1. The Partition of the Dataset.

IV. VISUALIZATION

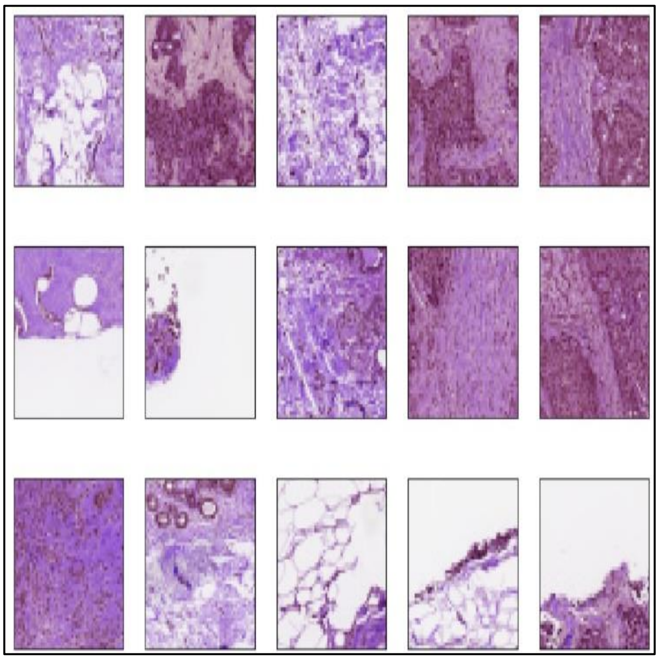

Fig:2 Breast Image Visualization

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR707

- Normal breast tissue should contain uniformly spaced cells and ducts, giving the appearance of having a normal structure.
- Breast tissue that is cancerous will probably contain cells that are clumped together or have strange shapes, giving the tissue a more erratic appearance.

V. MODELS USED FOR RICE DETECION

A. Convolution Neural Network:

There are many crucial steps in the Convolutional Neural Networks (CNNs) breast cancer detection process that are designed to efficiently analyse mammography pictures in order to spot any anomalies that could be signs of breast cancer. The CNN model is initially fed grayscale mammography pictures, which represent breast tissue. These photos are essential for identifying suspicious areas that can indicate the existence of malignant cells and provide the baseline information for further investigation.

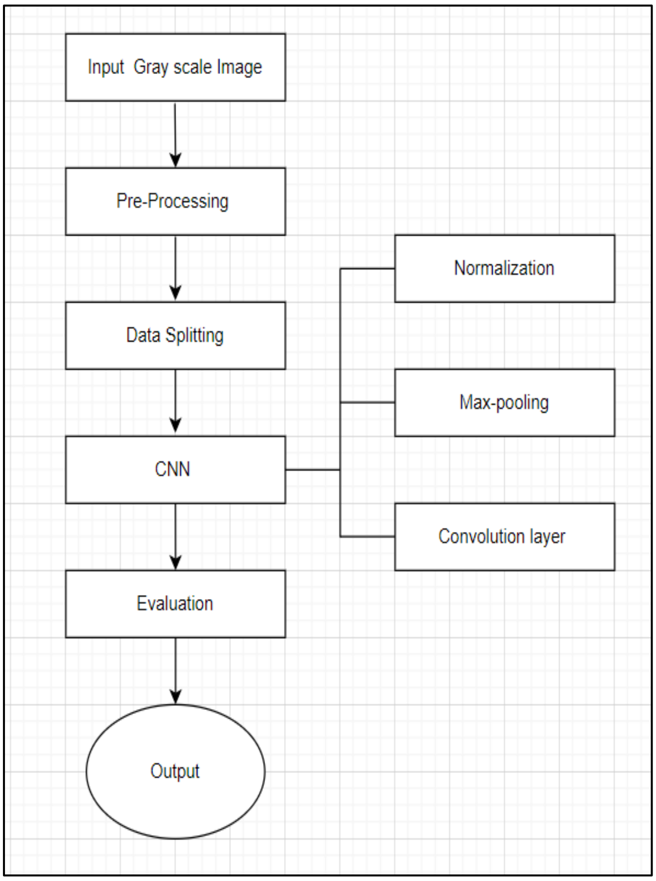

Fig. 3. Architecture for CNN

The pictures travel through several convolutional layers as they move through the CNN architecture. A variety of filters are applied by each convolutional layer in order to extract complex patterns, textures, and forms that are typical of breast cancer lesions. Through this procedure, the model is able to identify minute details in the mammography pictures, which helps with the accurate detection of spots that may be malignant.

To manage computational complexity and enhance model performance, pooling layers are incorporated between convolutional layers. These pooling layers reduce the spatial dimensions of the feature maps by down sampling them while maintaining essential features. By retaining crucial information and minimizing redundant data, these layers contribute to the overall efficiency of the CNN model in detecting breast cancer.

Following the convolutional and pooling layers, the flattened feature vectors are processed through fully connected layers. These layers serve as the high-level reasoning component of the model, where complex decision-making processes occur based on the extracted features. Here, the model learns to classify the input images into distinct categories, distinguishing between cancerous and non-cancerous instances with high accuracy. The model is rigorously evaluated employing different test or validation datasets after training. Performance metrics including accuracy, precision, recall, and F1 score are computed to assess how well the model performs in correctly classifying images of breast cancer. Ultimately, the trained CNN model might have real-world applications.

B. RESNET50V2

In the process of finding a set of methodical procedures designed to reliably detect malignant anomalies in mammography pictures. The RESNET50V2 model, which is well-known for its complex feature extraction and deep learning capabilities, uses these photos as its input data. They go through preprocessing in order to improve model performance and guarantee homogeneity. Resizing the photos to asimilar size, normalising the values of the pixel, and using augmentation methods to broaden the variety of the dataset are examples of preprocessing operations. can efficiently identify and extract pertinent information from the input photos, improving its precision in breast cancer detection.

It is made up of many deep leftover chunks. These blocks use several layers of convolutional and pooling procedures to let the model to learn complex patterns and features from the input data. improving its ability to distinguish between tiny indicators of breast cancer. High-level reasoning and decision-making processes take place in fully linked layers once the feature maps are flattened. The RESNET50V2 model can categorise the input photos into many categories, such malignant and non-cancerous, thanks to these layers. Using optimisation methods, the model modifies its parameters to reduce the discrepancy between the expected and actual results.

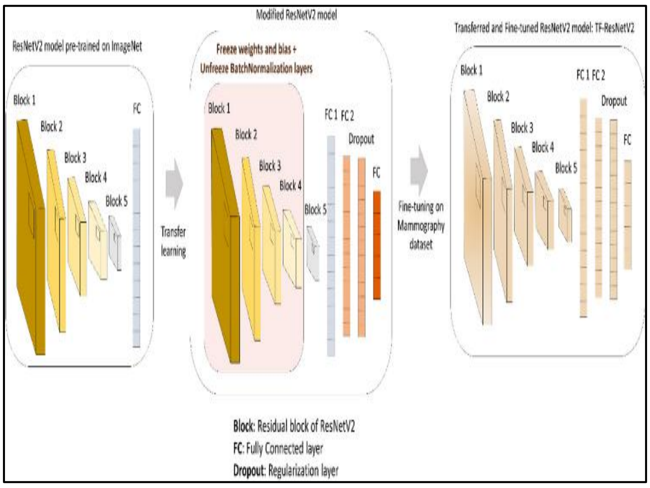

Fig:4 Transfer Learning for Breast Cancer Detection using a ResNetV2 Model.

- The ResNetV2 model—which was pre-trained on ImageNet—is being utilised to identify breast cancer in the context of the picture.
- The ResNetV2 model is displayed on the left of the picture. The model's convolutional layers are represented by the blocks with the labels "Block1" through "Block5".
- The layers labeled "FC 1" and "FC 2" are fully connected layers. These layers take the extracted features and learn to classify them as benign or malignant.

The process of optimising the ResNetV2 model for breast cancer detection is depicted on the image's right side. The model's biases and weights are fixed, with the exception of the Batch Normalization layers.

- The final layers of the model (FC 1 and FC 2) are unfrozen and trained.
- Only the last layers of the model are trained on the fresh dataset throughout the fine-tuning phase, with the majority of the model's weights frozen.

C.VGG16

There are numerous important phases in the detecting process. To improve the model's performance and standardise their format, mammography pictures are first preprocessed. This entails uniformly resizing the photographs, standardising pixel values, and using augmentation methods to broaden the variety of the collection. After the photos have been preprocessed, they are fed into the VGG16 architecture. This design's numerous layers allow the model to learn hierarchical representations of the input pictures, collecting both high- and low-level characteristics that are suggestive of anomalies related to breast cancer.

By use of extraction of the feature, the model is able to discern pertinent patterns from the photos, hence enhancing its capacity to recognise indicators of cancer. The model may then classify the input photos into different categories, such malignant or non-cancerous, by using the characteristics that were extracted. Metrics including The trained model's accuracy, precision, recall, and F1 score are produced to assess its efficacy in correctly categorizing breast cancer pictures using independent validation or test datasets.

Once trained and evaluated, the VGG16 model can be deployed for real-world breast cancer detection applications and which the prediction is completely accurate and correct in predicting the cancer.

VI. RESULTS

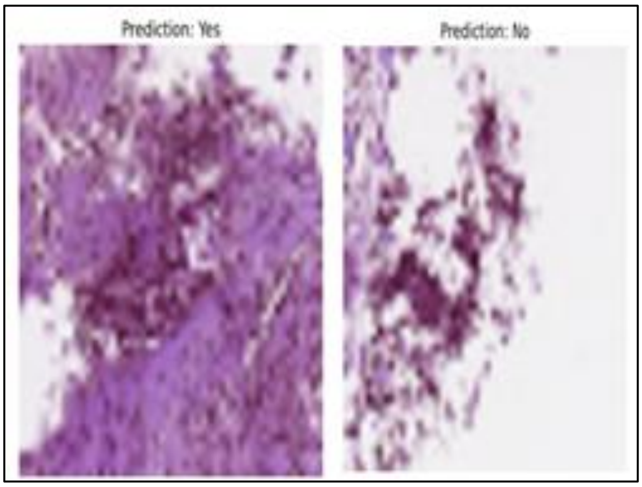

Fig:5 The Prediction of Cancer using Xray's

The above images describes in which the cancer is detected in which the image which shown in left is detected with cancer where as the image which is on right hand is no cancer images detected.

VII. CONCLUSION AND FUTURE WORKS

The ResNet50v2 model's 92% accuracy rate in detecting breast cancer highlights its usefulness for medical imaging applications. Subsequent endeavors may concentrate on augmenting the interpretability and generalizability of the model by the integration of supplementary clinical data and investigation of group techniques. Additionally, enhancing hyperparameters and pre-processing methods may help to increase accuracy and dependability. It would also be beneficial to use the model in actual clinical settings and carry out long-term research to evaluate its influence on patient outcomes. All things considered, more research and development might lead to advancements in machine learning-based breast cancer diagnosis and eventually better medical results.

REFERENCES

[1]. C. Dubey, N. Shukla, D. Kumar, A. K. Singh and V. K. Dwivedi, "Breast Cancer Modeling and Prediction Combining Machine Learning and Artificial Neural Network Approaches," 2022 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, 2022, pp. 119-124, doi: 10.1109/ICCCIS56430.2022.10037709.

[2]. A. Rovshenov and S. Peker, "Performance Comparison of Different Machine Learning Techniques for Early Prediction of Breast Cancer using Wisconsin Breast Cancer Dataset," 2022 3rd International Informatics and Software Engineering Conference (IISEC), Ankara, Turkey, 2022, pp. 1-6, doi: 10.1109/IISEC56263.2022.9998248.

[3]. A. Rovshenov and S. Peker, "Performance Comparison of Different Machine Learning Techniques for Early Prediction of Breast Cancer using Wisconsin Breast Cancer Dataset," 2022 3rd International Informatics and Software Engineering Conference (IISEC), Ankara, Turkey, 2022, pp. 1-6, doi: 10.1109/IISEC56263.2022.9998248.

[4]. M. Gupta and B. Gupta, "A Comparative Study of Breast Cancer Diagnosis Using Supervised Machine Learning Techniques," 2018 Second International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2018, pp. 997-1002, doi: 10.1109/ICCMC.2018.8487537.

[5]. E. A. Bayrak, P. Kırcı and T. Ensari, "Comparison of Machine Learning Methods for Breast Cancer Diagnosis," 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT), Istanbul, Turkey, 2019, pp. 1-3, doi: 10.1109/EBBT.2019.8741990.

[6]. N. Khuriwal and N. Mishra, "Breast cancer diagnosis using adaptive voting ensemble machine learning algorithm," 2018 IEEMA Engineer Infinite Confere

[7]. A. Bah and M. Davud, "Analysis of Breast Cancer Classification with Machine Learning based Algorithms," 2022 2nd International Conference on Computing and Machine Intelligence (ICMI), Istanbul, Turkey, 2022, pp. 1-4, doi: 10.1109/ICMI55296.2022.9873696.

[8]. Anshuman and U. Kumar, "Machine Learning model for detection of Breast Cancer," 2021 5th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, 2021, pp. 1-4, doi: 10.1109/ISCON52037.2021.9702416.

[9]. M. Akhil and P. V. S. Kumar, "Breast Cancer Prognosis using Machine Learning Applications," 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 2022, pp. 488-493, doi: 10.1109/ICAC3N56670.2022.10074517.

[10]. V. A. Telsang and K. Hegde, "Breast Cancer Prediction Analysis using Machine Learning Algorithms," 2020 International Conference on Communication, Computing and Industry 4.0 (C2I4), Bangalore, India, 2020, pp. 1-5, doi: 10.1109/C2I451079.2020.9368911.

[11]. Y. Tewari, E. Ujjwal and L. Kumar, "Breast Cancer Classification Using Machine Learning," 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2022, pp. 01-04, doi: 10.1109/ICACITE53722.2022.9823932.

[12]. G. Neelima, P. Kanchanamala, A. Misra and R. A. Nugraha, "Detection Of Breast Cancer Based on Fuzzy Logic," 2023 International Conference on Advancement in Data Science, E-learning and Information System (ICADEIS), Bali, Indonesia, 2023, pp. 1-6, doi: 10.1109/ICADEIS58666.2023.10270874.

[13]. P. Divya, D. P. Rajan, R. Suguna and S. Velliangiri, "Machine Learning Techniques for Prediction and Analysis of Benign and Malignant in Breast Cancer," 2022 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2022, pp. 1-6, doi: 10.1109/ICCCI54379.2022.9740814.

[14]. A. Kajala and V. K. Jain, "Diagnosis of Breast Cancer using Machine Learning Algorithms-A Review," 2020 International Conference on Emerging Trends in Communication, Control and Computing (ICONC3), Lakshmangarh, India, 2020, pp. 1-5, doi: 10.1109/ICONC345789.2020.9117320.

[15]. R. H. Khan, J. Miah, M. M. Rahman and M. Tayaba, "A Comparative Study of Machine Learning Algorithms for Detecting Breast Cancer," 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 2023, pp. 647-652, doi: 10.1109/CCWC57344.2023.10099106.

[16]. https://drive.google.com/drive/folders/1LBFVmadDSORb5aLdSGlvlg15RnXtmzdQ