iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** II **Month of publication:** February 2024

DOI: https://doi.org/10.22214/ijraset.2024.58469

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue II Feb 2024- Available at www.ijraset.com

Breast Cancer Detection Using Artificial Intelligence

Mr. Amratsingh Johar[1], Ms. Bhairavi Kelwadkar[2], Dr. Ashutosh Lanjewar[3]
[1, 2]*Department of Computer Science and Engineering, J D College of Engineering & Management, Nagpur*
[3]*Department of Artificial Intelligence, J D College of Engineering & Management, Nagpur*

Abstract: ***Breast cancer is one of the most common diseases faced by women and causes more deaths every day. It is considered one of the greatest diseases in medical history. Automated diagnostic systems not only assist diagnosticians but also provide reliable, efficient and fast services that reduce the likelihood of death. This paper proposes a model to manually classify breast cancer samples into benign and malignant subcategories using multiple algorithms to predict interpretation. It is generally considered as all data containing a set of 30 characteristics related to the nuclei found in digitized images of fine needle aspiration (FNA) of the breast. All acquired features are visualized and understood to gain insight for future diagnoses. Breast cancer prediction uses support vector distribution (SVC) in its ability to identify new, unrecognized data based on a set of features. The application of principal component analysis (PCA) dimensionality reduction strategy complements the applicability of features to solve the eigenvector problem. Final results were visualized using conflict matrices and receiver operating characteristic curves (ROC). The first evaluation compares six popular algorithms, and the K-nearest Neighbor (KNN) algorithm starts to show promise, but in the end the support vector machine with optimized hyperparameters appears to be the most accurate model. This paper also presents a deep learning based breast cancer diagnosis using a multilayer convolutional neural network (CNN). The model is based on the ResNet50V2 architecture and has been proven to achieve state-of-the-art results on a variety of image classifications. The model consists of convolutional, layered and full (dense) layers. Convolutional techniques learn to extract features from input images, while pooling techniques reduce the dimensionality of the feature map. The full convolution method combines the features extracted by the convolution method and classifies the image as cancerous or non-cancerous. The model was trained on a breast histopathology imaging dataset and achieved an accuracy rate of 95%. This shows that the proposed model can detect breast cancer from histopathology images. This paper also shows how to use the learned model to predict whether the image contains signs of breast cancer. This can be done by loading the model and then passing the image as input to the model. The model will assign an outcome score to each category (cancer and non-cancer). If the cancer score is greater than the threshold, the image is predicted to be cancerous. Otherwise, it is estimated that the image is not cancerous.***

Keywords: Artificial Neural Network (ANN), K-Nearest Neighbor (KNN), Fine Needle Aspiration (FNA), Convolutional Neural Network (CNN), Receiver Operating Characteristic Curve (ROC), Principal Component Analysis (PCA), Support Vector Classification (SVC).

I. INTRODUCTION

Breast cancer casts a long shadow, claiming lives and demanding swift, accurate detection. Early diagnosis unlocks a world of better treatment options and higher survival rates. But traditional methods often falter, limited by human error and scalability. This is where advanced tools like machine learning and deep learning step in, offering promising solutions. This paper explores two such tools, each wielding their unique strengths against this formidable foe. The first, a machine learning model, meticulously analyzes 30 features extracted from cell nuclei in fine needle aspirate images. Using Support Vector Classification, it discerns benign and malignant breast cancer subtypes, paving the way for accurate diagnosis. Techniques like Principal Component Analysis further refine its feature analysis, sharpening its capabilities. The second tool, a deep learning model, harnesses the power of a Convolutional Neural Network built on the ResNet50V2 architecture. This model delves into the intricate world of histopathology images, extracting features through convolutional and pooling layers. Then, using its fully-connected layers, it classifies images as cancerous or non-cancerous. Trained on a vast dataset, this model exhibits remarkable accuracy, offering a powerful weapon in the fight against cancer. But this is just the tip of the iceberg. This study delves deeper, exploring the use of machine learning pipelines to automate and streamline the development of these models. By leveraging scikit-learn's pipeline functionality, we were able to seamlessly integrate data standardization and model training into a single workflow.

©IJRASET: All Rights are Reserved

This not only sped up the development process but also enabled efficient evaluation of multiple algorithms, paving the way for faster, more robust breast cancer detection models in the future. Our work underscores the immense potential of machine learning and deep learning in automating and enhancing breast cancer detection. By embracing these innovative tools and streamlined workflows, we can open doors to improved diagnosis, ultimately leading to better care for patients battling this disease. It highlights the benefits of using machine learning pipelines for efficient evaluation of algorithms and paves the way for future advancements in breast cancer detection.

II. LITERATURE REVIEW

Early detection is key to beating breast cancer, and machine learning offers powerful tools for improvement. Traditional methods like SVMs excel in analyzing Fine-Needle Aspiration (FNA) data, while ensemble methods further boost accuracy. Deep learning, especially Convolutional Neural Networks (CNNs) like ResNet50V2, shines in analyzing histopathology images, achieving remarkable results. However, challenges remain. Data access, quality, and diversity are crucial, but limitations exist. Deep learning models can be opaque, hindering clinical adoption. Additionally, bias in data and algorithms can lead to unfair outcomes, necessitating ethical development and deployment. Despite these challenges, the future is promising. Integrating clinical data with images holds immense potential, while explainable AI techniques are being developed to demystify deep learning models. Data privacy and security concerns require careful attention, but the potential for improved breast cancer detection using these techniques is undeniable. This review provides a starting point, but for deeper understanding, consult original research and expert opinions on the latest advancements in this rapidly evolving field.

III. PROBLEM STATEMENT

AI advances in breast cancer detection, critical hurdles persist. Limited accuracy across diverse patients, data scarcity hindering robust models, "black box" models raising ethical concerns, and integration challenges with clinical workflows and regulations pose significant obstacles. Ensuring equitable access across social and economic backgrounds further complicates the picture. Collaborative efforts to tackle these challenges are crucial to unlock AI's true potential for improved detection, personalized care, and ultimately, saving lives.

IV. METHODOLOGY

Machine learning algorithms and deep learning techniques make up the few key components of the suggested strategy for predicting and detecting of breast cancer. In this study, 569 samples of malignant and benign tumour cells were collected from Fine Needle Aspiration(FNA) dataset which predicts breast cancer outcome and a well maintained dataset of breast histopathological images with variety of breast cancer patient cases. The testing stage was applied to many individuals, whether they had breast cancer or not. Several models and techniques were put forth to improve the accuracy of Breast Cancer diagnosis. Logistic Regression (LR), Artificial Neural Network (ANN), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Classification and Regression Tree (CART), Naive Bayes (NB), and Linear Discriminant Analysis (LDA) are a few of the techniques used. Deep learning techniques, such as neural networks (CNN) is also used to evaluate medical images (mammograms, MRIs, Histopathological images and so on) and discover cancer patterns. To assure accurate cancer detection and reduce false positives/negatives, first construct models with high accuracy, sensitivity and specificity. Create models for early cancer detection to allow for timely intervention and improved patient outcomes. Implemented a binary classification model leveraging state-of-the-art techniques such as Convolutional Neural Networks (CNN) and CNN Classifiers. Utilize a pre-trained ResNet50V2 architecture for efficient feature extraction from medical images. Train the model on a dataset specifically designed for breast cancer detection. Evaluate the deep learning model's ability to process complex patterns in the data.

A. Exploratory Data Analysis

Before diving into models, exploratory data analysis (EDA) sheds light on the data's true nature. This crucial step involves both numerical summaries (descriptive statistics) and visual exploration (unimodal and multimodal approaches). Described by 357 cancer-free observations and 212 cancerous ones, the data hints at potential exponential distributions for attributes like perimeter, radius, area, concavity, and compactness. Visualizations further reveal strong positive correlations between mean values within certain parameters (1 to 0.75 range). By understanding these initial insights gained a powerful foundation for building effective models to tackle breast cancer.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

B. Data Pre-Processing

Before diving into the intricate world of machine learning, a crucial step paves the way: data preprocessing. This meticulous process transforms raw data into a polished canvas, revealing its hidden structure and preparing it for the algorithms' discerning eyes. Missing values, those pesky gaps, are carefully filled, ensuring data integrity. Outliers, the data's oddballs, are identified and removed. Feature selection techniques, like PCA, work their magic, condensing the data's complexity into a manageable 2-dimensional space. This allows algorithms to focus on the most relevant features, maximizing their learning potential. Finally, the data is split into training and testing sets. The training set becomes the algorithms' playground, where they learn the data's secrets. The testing set, the ultimate challenge, evaluates their newfound knowledge. Feature standardization ensures fair play, with all features scaled to a common ground. This prevents biases and allows algorithms to focus on the true relationships within the data and obtaining the Eigenvectors and Eigenvalues from the covariance matrix or correlation matrix. By traversing these steps, we transform raw data into a powerful tool, ready to unlock the full potential of machine learning in the fight against breast cancer.

C. Predictive Model Using Support Vector Machine (SVM)

SVMs excel at handling nonlinear data through kernel transformations, making them well-suited for cancer diagnosis where feature relationships are often complex. They define precise decision boundaries even with few features, but require careful preprocessing and parameter tuning. While powerful, they can be less interpretable than tree-based models. Key parameters include the regularization parameter (balancing model complexity and generalization), kernel function (mapping data to higher dimensions), and gamma and C (controlling model complexity). Cross-validation, specifically 3-fold, ensured model generalizability and prevented overfitting. Performance evaluation involved ROC curves, AUC metrics, and confusion matrices, revealing strengths and weaknesses in accuracy, sensitivity, specificity, precision, and prevalence.

D. Optimizing With The SVM Classifier

While machine learning models hold remarkable promise for early breast cancer detection, unlocking their full potential demands a careful dance between model architecture and fine-tuning. Like a master conductor guiding an orchestra, we meticulously adjust model hyperparameters akin to tuning individual instruments to achieve a harmonious symphony of accuracy, precision, and recall. To navigate this intricate tuning process, we leverage techniques such as 10-fold cross-validation and grid search, systematically exploring various combinations of kernel types, regularization parameters, and feature selections. Data standardization ensures all features play on a level field, enhancing model clarity and performance. Advanced optimization algorithms, such as random search, Bayesian optimization join the ensemble, offering diverse strategies to discover the most optimal hyperparameter configuration. We visualize decision boundaries to unveil the model's classification patterns, ensuring unbiased and reliable predictions. We address potential class imbalances and explore ensemble methods that strategically combine multiple models, potentially amplifying model precision. By diligently studying successful SVM optimization research in breast cancer prediction and tailoring these techniques to our unique dataset, we forge a powerful alliance between human expertise and machine learning's capabilities. This synergistic approach offers a promising path toward earlier detection, improved outcomes, and ultimately, a brighter future for those facing breast cancer.

E. Automating Breast Cancer Detection With Machine Learning Pipelines

This study explores the use of machine learning pipelines to automate and streamline the development of breast cancer detection models. By leveraging scikit-learn's pipeline functionality, we were able to seamlessly integrate data standardization and model training into a single workflow, enabling efficient evaluation of multiple algorithms. Our initial evaluation included six algorithms: Logistic Regression, Linear Discriminant Analysis, K-Nearest Neighbors, Decision Tree, Naive Bayes, and Support Vector Machine (SVM). While K-Nearest Neighbors yielded promising results, data standardization significantly enhanced SVM performance. To further optimize model performance, we employed GridSearchCV for hyperparameter tuning of both SVM and K-Nearest Neighbors. This process identified optimal configurations for each algorithm, leading to improved accuracy.

F. Diagnosis And Classification

Classification is the process of dividing a set of data into categories. This can be done on structured and unstructured data. Class prediction of the provided data points is the first step in the process. The main purpose is to determine to which category or class the new data belongs. In our research on AI-based breast cancer diagnosis, the classification proved useful in classifying histopathology images as cancerous or non-cancerous, as shown in Fig. 1.

©IJRASET: All Rights are Reserved

We harness the power of convolutional neural networks (CNNs), known for their ability to extract hidden patterns from complex images. In addition to CNN, we tested and compared various algorithms such as logistic regression, K-nearest neighbors, and support vector machines to determine how unique our data was. This comprehensive evaluation ensures that the most accurate and data-appropriate algorithm is selected. Through rigorous analysis and search algorithm, we focus on intelligence-based best breast cancer diagnosis, which can improve patient accuracy and result effect.

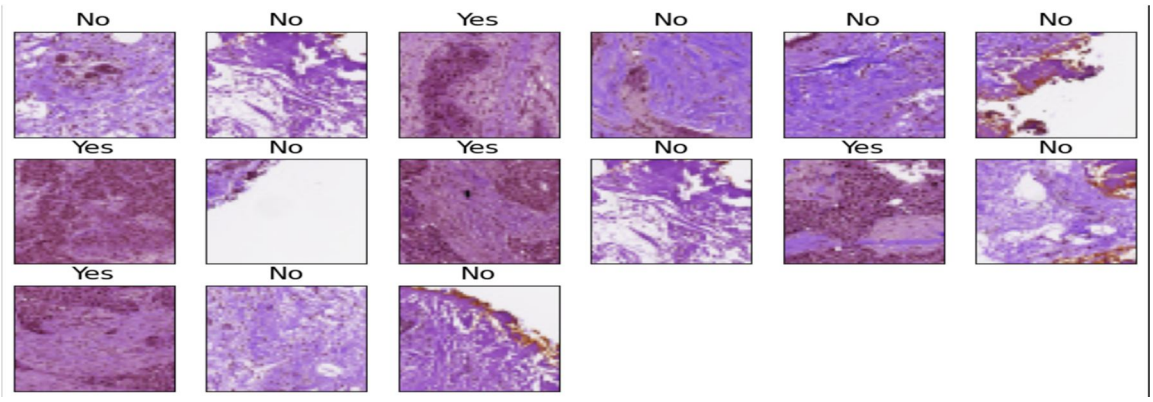


Fig. 1 Glimpse of the images in the dataset

Fig. 1 represents the visualization which provides a quick glimpse of the images in the dataset, helping to understand the types of data working with and the distribution of tumor and non-tumor cases.

G. Defining The Base Model Architecture

For AI cancer diagnosis,a powerful ResNet50V2 is used, a deep neural network pre-trained on millions of images. This prior knowledge gives us an advantage, allowing us to customize the model for a specific project with less information and training time. We kept ResNet50V2's core in good shape and kept its weight low to keep its performance first in the main features. On top of that, we added some specific methods to translate this into a classification of breast cancer cells. We compress the image data, apply the output to prevent overfitting, and finally add a layer to predict the presence of cancer with a simple yes or no urine. ResNet50V2's deep architecture is effective in extracting pattern data from complex histopathological images; this is an important skill for accurate tissue identification. By combining the knowledge gained before training with our modifications, we aim to create a powerful intelligence that can revolutionize breast cancer treatment.

V. FUTURE SCOPE

While the aforementioned AI-driven approaches offer a glimpse into a future with enhanced breast cancer detection, the vast potential of AI remains largely untapped. Future Scope of AI in Breast Cancer Detection:

1) *Expand AI Applications:* Predict individual risk, treatment response, and discover new drugs. Integrate AI with genomics, nanotech for more comprehensive approaches.
2) *Improve Data Quality & Accessibility:* Create large, diverse datasets through collaboration. Address privacy and security concerns.
3) *Ensure Ethical Considerations & Transparency:* Guarantee fairness, accountability, and responsible AI use. Develop interpretable models for better understanding.
4) *Accelerate Clinical Translation:* Conduct rigorous validation studies and establish regulations. Train healthcare professionals for effective AI utilization.
5) *Explore Emerging Technologies:* Leverage quantum computing for faster analysis and blockchain for secure data sharing.
6) *Address Social & Economic Equity:* Implement strategies for equitable access across populations. Combat potential biases in data and algorithms.
7) *Consider Psychological Impact:* Educate and support patients to address anxiety and concerns. Develop user-friendly interfaces and communication strategies.
8) *Ensure Continuous Evaluation & Improvement*: Monitor and evaluate AI systems for effectiveness, safety, and ethics. Foster a culture of ongoing learning and improvement.

©IJRASET: All Rights are Reserved

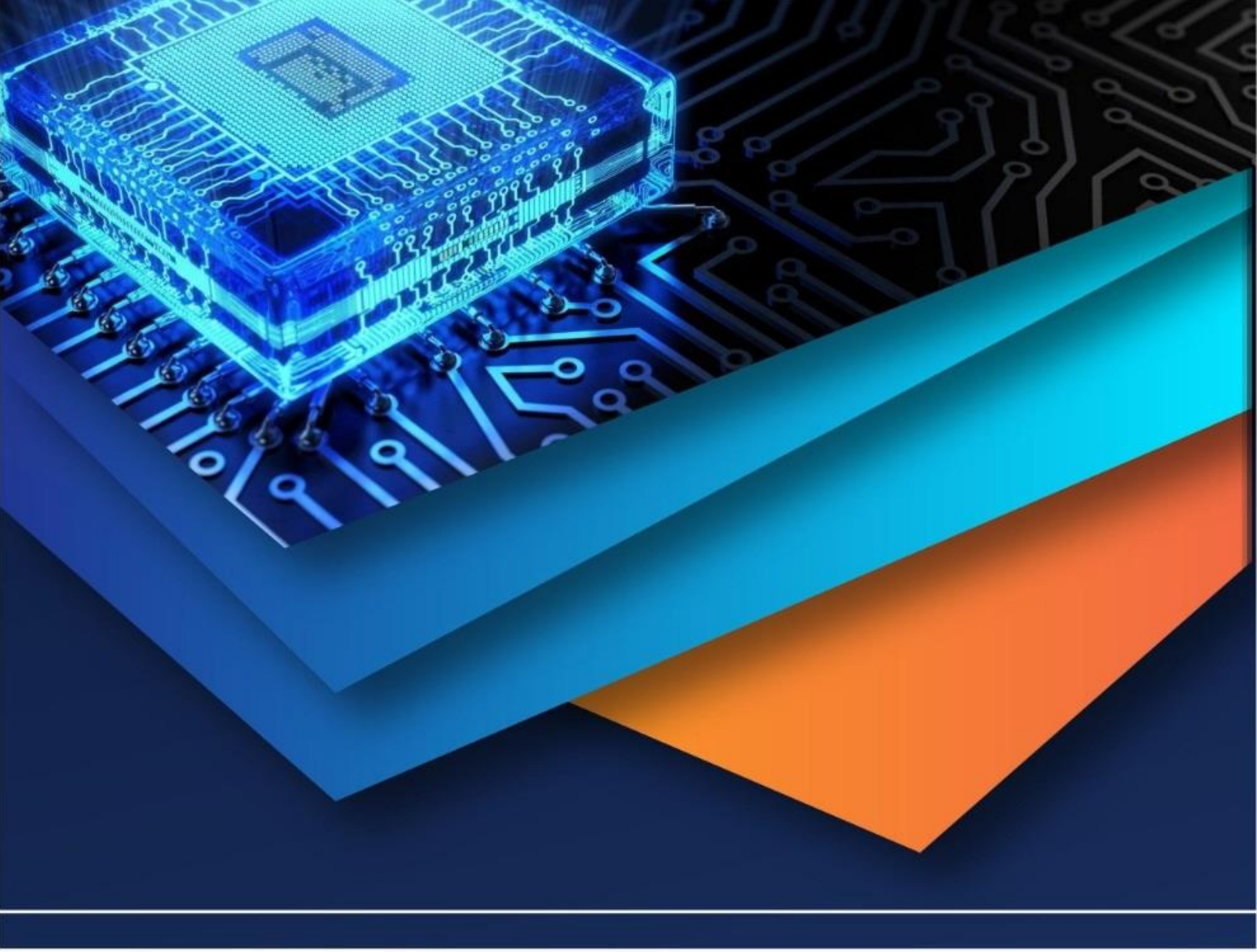
VI. CONCLUSION

This study explored the potential of AI in breast cancer detection through a combined approach utilizing both FNA data and histopathological image analysis. In the prediction stage, various supervised learning techniques, including SVMs and KNNs, achieved high accuracy, suggesting their potential for non-invasive risk assessment. In the detection stage, the ResNet50V2 deep learning architecture demonstrated promising performance in analyzing histopathological images for cancer detection. These findings highlight the potential benefits of AI in breast cancer diagnosis and management. Combining FNA and histopathological data with AI offers a multifaceted approach, providing valuable insights for both early detection and definitive diagnosis. The high accuracy achieved by traditional machine learning algorithms in prediction suggests their potential for clinical application, while the promising performance of ResNet50V2 demonstrates the potential of deep learning in image-based diagnosis. By addressing these challenges and leveraging the combined power of FNA and histopathological data, AI has the potential to revolutionize breast cancer detection, ultimately leading to improved patient outcomes and potentially saving lives.

REFERENCES

[1] Ali Bou Nassif, Manar Abu Talib, Qassim Nasir, Yaman Afadar, Omar Elgendy, “Breast cancer detection using artificial intelligence techniques: A systematic literature review,” Artificial Intelligence in Medicine, Elsevier, Vol 127, May 2022.
[2] Md Haris Uddin Sharif, “Breast Cancer Detection using Artificial Neural Networks,” International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429 Volume 9 Issue X Oct 2021
[3] Alok Chauhan, Harshwardhan Kharpate, Yogesh Narekar, Sakshi Gulhane, Tanvi Virulkar,Yamini Hedau,“Breast Cancer Detection and Prediction using Machine Learning,” 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA) 02-04 September 2021.
[4] Sarthak Vyas, Abhinav Chauhan, Deepak Rana Mohd, Noman Ansari, “Breast Cancer Detection Using Machine Learning Techniques,” International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 10 Issue V May 2022.
[5] “Breast Cancer Diagnosis in Digital Breast Tomosynthesis: Effects of Training Sample Size on Multi-Stage Transfer Learning Using Deep Neural Nets, IEEE Journals & Magazine, IEEE Xplore,” ieeexplore.ieee.org. Jan. 23, 2024.
[6] Maged Nasser and Umi Kalsom Yusof “Deep Learning Based Methods for Breast Cancer Diagnosis: A Systematic Review and Future Direction”, Artificial Intelligence in Medicine, Elsevier, Vol 127, May 2022.
[7] E. Mahoro and M. A. Akhloufi, “Applying Deep Learning for Breast Cancer Detection in Radiology,” Current Oncology, vol. 29, no. 11, pp. 8767–8793, Nov.2022,doi:https://doi.org/10.3390/curroncol291 0690.
[8] Youusif A. Hamad, Konstantin Simonov, Mohammad B. Naeem “Breast Cancer Detection and Classification Using Artificial Neural Networks”, 1st Annual International Conference on Information and Sciences (AiCIS) November 2018.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)