IJCRT.ORG ISSN : 2320-2882
INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

BREAST CANCER DETECTION USING DEEP LEARNING

"Revolutionizing Breast Cancer Detection: A Deep Learning Approach"

[1]Dr. Abhijit D. Jadhav, [2]Ajinkya Raje, [3]Tejas Chakkarwar, [4]Rajnandini Jagtap, [5]Chetana Thorat
[1]Professor, [2]Student, [3]Student, [4]Student, [5]Student
[1]Department Of Computer Engineering,
[1]PCET's Pimpri Chinchwad College Of Engineering and Research, Pune, India

Abstract: Breast cancer is the second most common cause of mortality for women globally. However, with early identification and prevention, the risk of dying can be considerably reduced. Breast cells are the site of development for breast cancer, which affects women rather frequently. Breast cancer is the disease that claims the lives of the most women after lung cancer. In existing systems which use algorithms such as Decision Trees, SVM, Random Forest etc. have limitations like scalability, overfitting of the data, feature dependence, computationally intensive etc. which we try to overcome by using Deep Neural Networks like CNN. In order to improve automated breast cancer identification (WSI), we propose a convolutional neural network (CNN) technique in this study by analyzing hostile ductal carcinoma tissue zones in whole-slide pictures. This overview looks at the results of a proposed system for automatically diagnosing breast cancer using several convolutional neural network (CNN) designs and compares them to those achieved using machine learning (ML) techniques. All buildings were built from a large database.Validation tests were conducted in order to provide quantifiable results using the performance parameters of each approach. Functions like VGG16 and ResNet50 are used in this model to improve the accuracy of the model. Both architectures have shown good performance in image classification tasks, including breast cancer detection. In the last layer of architecture Softmax Activation function is used with which we can obtain meaningful class probabilities for breast cancer prediction. This discusses numerous statistics and examines breast cancer datasets to improve the accuracy of breast cancer diagnosis using convolutional neural network techniques. After analysis, the data may be looked at, evaluated, and used for training. Following that, error histograms were extracted from the dataset in order to create the confusion matrix. We may estimate accuracy levels as a result, and we can achieve high accuracy above 90 percent. The outcomes of a breast cancer prediction model can differ based on various factors, including the dataset's quality and size, the selected model architecture, and the employed optimization techniques. Ultimately, the prediction model yields a classification result that indicates whether an individual has breast cancer or not.

***Index Terms* - CNN, Big data, Healthcare data, Machine Learning, Image processing, ResNet50, VGG16.**

I. INTRODUCTION

Breast cancer is the second most common kind of cancer in both men and women throughout the world. In 2012, it contributed to about 12% of all new cancer cases and 25% of all cancers in women. Breast cancer's beginning causes breast cells to begin to multiply uncontrolled. Tumors caused by these cells regularly form and can be seen on x-rays and felt as bumps. If a tumor's cells have the capacity to infiltrate neighboring tissues or spread to other regions of the body, it is considered malignant (cancerous).

Breast cancer can be successfully treated if discovered early. It is therefore essential to have readily available screening methods for detecting breast cancer's early signs and symptoms. Several imaging techniques are employed for this disease screening. Common techniques include mammography, ultrasound, and thermography. One of the most important methods for the early detection of breast cancer is mammography. Ultrasonography is commonly used since mammography is worthless for tight breasts. Given these worries, tiny tumors can be avoided by radiographic radiation, and thermography may be more effective than ultrasound for detecting smaller malignant masses.

Decision trees are a popular machine learning technique for classification tasks. They construct a tree-like model of decisions and their possible consequences. Decision trees are intuitive and can handle both categorical and numerical data. However, they tend to overfit the training data, leading to poor generalization on unseen data. Additionally, decision trees may struggle with capturing complex relationships between features.
Naive Bayes is computationally efficient and can handle high-dimensional data well, but the assumption of feature independence may not hold in some cases, limiting the model's performance. Similarly, Random Forest Algorithm may also suffer from overfitting if the number of trees is not properly tuned and Logistic Regression assumes a linear relationship between the variables, which may not capture complex interactions or non-linear patterns in the data.

Deep Learning overcomes many of these limitations mentioned earlier, that’s why it has gained popularity and demonstrated superior performance in various domains, including medical image analysis and breast cancer prediction. CNNs excel at capturing intricate and non-linear patterns in data. In the case of breast cancer prediction, CNNs can automatically learn discriminative features from medical images, such as mammograms or histopathology slides. This allows the model to identify subtle abnormalities and patterns that may not be easily detected by human observers or traditional machine learning algorithms. Scalability is one of the other features which is present here, Deep learning models including CNNs, can scale effectively to large datasets. CNNs enable end-to-end learning, where the model learns the optimal features and performs the classification task simultaneously. Traditional machine learning approaches often require manual feature engineering, where domain experts manually design and extract relevant features from the data. CNNs are inherently robust to variations in image appearance, such as changes in lighting conditions, image resolution, or noise. They can generalize well to unseen data and are less sensitive to minor variations that may occur in medical images.

II. PROPOSED TECHNIQUE

In this article, we propose a convolutional neural network (CNN) technique by analyzing hostile ductal carcinoma tissue zones in whole-slide images to enhance automatic breast cancer identification (WSI).
This overview examines the outcomes of multiple convolutional neural network (CNN) architectures used in a proposed system for autonomously identifying breast cancer and contrasts them to the outcomes obtained using machine learning (ML) methods. All structures were created using a sizable database.
To get quantifiable findings using the performance characteristics of each strategy, validation tests were carried out. This model uses features like VGG16 and ResNet50 to increase model accuracy.
The specific results achieved by a breast cancer prediction model can vary depending on factors such as the quality and size of the dataset, the choice of the model architecture, and the optimization techniques used. The result which we obtain at the end of the prediction is a classification result which tells if someone has breast cancer or not.

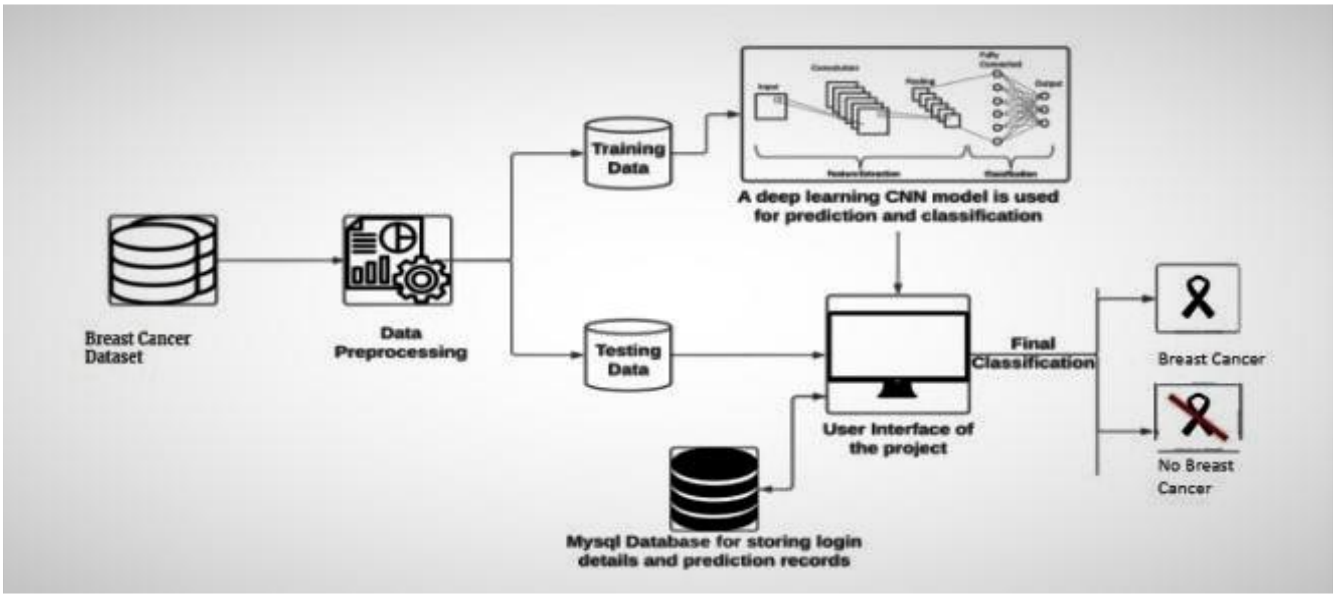

figure -proposed technique

III. Challenges

Predicting breast cancer prediction using deep learning techniques can be challenging due to various factors. Here are some of the key challenges associated with this task:

1. Insufficient and imbalanced data: Deep learning models require large amounts of high-quality data for effective training. However, obtaining a diverse and well-annotated dataset for breast cancer prediction can be challenging. Moreover, imbalanced class distributions, where the number of positive (cancer) cases is significantly smaller than negative (non-cancer) cases, can lead to biased models.

2. Interpretability and transparency: Deep learning models are often considered black boxes, meaning they lack interpretability and transparency in their decision-making process. Understanding the reasoning behind the model's predictions is crucial in medical applications, as it helps clinicians trust and validate the predictions.

3. Overfitting and generalization: Deep learning models are prone to overfitting, especially when dealing with limited data. Overfitting occurs when the model performs well on the training data but fails to generalize to unseen data. Robust techniques, such as regularization, cross-validation, and data augmentation, need to be employed to mitigate overfitting and improve generalization.

4. Limited availability of annotated medical data: Annotated medical datasets are scarce and often require expert knowledge for accurate labeling. Collecting and labeling large-scale datasets for breast cancer prediction can be time-consuming and resource intensive.

5. Handling multimodal data: Breast cancer prediction may involve different types of data, such as mammograms, clinical records, genetic information, and pathology reports. Developing effective deep learning models that can seamlessly integrate and learn from these multimodal data sources is a complex task.

6. Ethical considerations and biases: Deep learning models are susceptible to biases present in the data used for training. Biases in medical data, such as racial or socioeconomic biases, can lead to disparities in the accuracy and fairness of breast cancer predictions. Careful consideration of ethical implications and fair representation in dataset construction is necessary to mitigate these biases.

7. Model robustness and uncertainty estimation: Deep learning models are sensitive to variations and adversarial attacks. Ensuring robustness against noise, artifacts, or perturbations in the input data is crucial for reliable breast cancer predictions. Additionally, estimating model uncertainty and providing confidence intervals for predictions are essential in clinical decision-making.

Addressing these challenges requires interdisciplinary collaboration between deep learning researchers, medical professionals, and domain experts. Striving for transparency, interpretability, and fairness in model development is crucial to build trust and facilitate the adoption of deep learning in breast cancer prediction.

IV. METHODOLOGIES

The stages involved in the methods for problem-solving in breast cancer diagnosis include data preparation, feature extraction, model training, and assessment, to name just a few. By assembling a variety of high-quality datasets, identifying important traits, and training, healthcare professionals can develop precise methods for early detection. These models are evaluated using performance metrics and validated with independent datasets to ensure that they are successful in correctly categorizing breast cancer cases and ultimately improving patient outcomes.

The VGG16 and ResNet50 functions will be used in this breast cancer prediction model to make predictions. In both situations, the decision of which model to use—VGG16 or ResNet50—depends on a number of variables, including the size of the dataset, the available computing power, and the particulars of the breast cancer prediction task. In image classification tasks, such as the detection of breast cancer, both designs have demonstrated strong performance. To find the optimal solution for a given issue, it's critical to experiment with several models and methodologies. Both the VGG16 and ResNet50 pretrained models were developed using sizable image datasets.

In this, we simply train the succeeding layers for breast cancer prediction and freeze the early layers of ResNet50 and VGG16. And for adjusting it to the breast cancer prediction task, add extra layers on top of ResNet50 and VGG16. We apply the softmax activation function in the last layer of the VGG16 model and ResNet50 model because doing so allows us to get useful class probabilities for breast cancer prediction. Because of the learnt features and patterns in the data, this enables us to make certain predictions and evaluate the likelihood that an image belongs to distinct groups.

V. FUTURE SCOPE

The future scope of making breast cancer prediction using deep learning techniques is promising, with several potential advancements and opportunities:

1. Enhanced accuracy and early detection: Deep learning models have the potential to improve the accuracy and early detection of breast cancer. As more data becomes available, including multimodal data like genetic information and patient history, deep learning algorithms can learn complex patterns and features that may aid in early diagnosis and personalized treatment recommendations.

2. Integration with imaging technologies: Deep learning models can be integrated with advanced imaging technologies, such as digital mammography, digital breast tomosynthesis, and magnetic resonance imaging (MRI). These models can assist radiologists in detecting subtle abnormalities, segmenting tumors, and characterizing breast lesions more accurately, leading to improved diagnostic accuracy.

3. Predictive models for treatment response: Deep learning can be utilized to predict individual patient response to different treatment options. By leveraging large-scale datasets that include clinical information and treatment outcomes, deep learning models can identify patterns that contribute to treatment response and provide personalized treatment recommendations.

4. Decision support systems: Deep learning algorithms can serve as decision support systems for clinicians, assisting them in interpreting complex data and making informed decisions. By integrating patient-specific information, such as medical history, genetic profiles, and imaging results, these systems can provide more accurate risk assessments and treatment plans.

5. Integration of genomics and proteomics data: The availability of genomic and proteomic data has opened up new avenues for understanding the underlying molecular mechanisms of breast cancer. Deep learning models can analyze and integrate these complex molecular datasets to identify biomarkers, classify subtypes, and predict patient outcomes.

6. Transfer learning and model generalization: Transfer learning, where pre-trained deep learning models are fine-tuned on specific breast cancer datasets, can help overcome the challenge of limited labeled data. By leveraging knowledge learned from related tasks or domains, transfer learning can enhance model generalization and improve performance on smaller datasets.

7. Explainable deep learning models: Addressing the interpretability challenge of deep learning models is crucial for their wider acceptance in clinical settings. Future research may focus on developing explainable deep learning models that can provide clinicians with insights into the reasoning behind predictions, helping them trust and validate the model's outputs.

8. Integration with electronic health records (EHR): Deep learning models can be integrated with electronic health records to leverage longitudinal patient data for better breast cancer prediction and management. By analyzing a patient's complete medical history, including previous diagnoses, treatments, and outcomes, deep learning models can provide more comprehensive risk assessments and personalized recommendations.

VI. Conclusion

To summarize, this study showcases the effectiveness of deep learning methods, specifically VGG16 and ResNet50 networks, in classifying normal and atypical breast cancer. The performance of the classification algorithms was evaluated using precision, recall, and accuracy rate metrics. The results demonstrate that VGG16 achieved the highest accuracy score of 94 percent, indicating its ability to accurately differentiate between the two types of breast cancer. The implications of this research are significant for the medical field as accurate classification of breast cancer is crucial for determining appropriate treatment plans and improving patient outcomes. The findings suggest that deep learning algorithms, especially VGG16, can enhance the diagnostic process by improving the accuracy of breast cancer classification.

VII. Acknowledgement

We would like to express our gratitude to our principal - Dr .H.U Tiwari and HOD - Dr. Archana Chaughule and our guide Dr. Abhijeet D. Jadhav who were a continual source of inspiration, for being of great support and guiding us through the research. Their extensive knowledge, experience and expertise enabled us to successfully complete this project. This effort would not have been possible without their help and supervision. This initiative would not be successful without the contribution of everyone. We were always there to encourage each other and that kept us together until the end.

References

[1] Essam H. Houssein1, Marwa M. Emam1, Abdelmgeid A. Ali1(2021) Neural Computing and Applications https://doi.org/10.1007/s00521-022-07445-5(0123456789().,- volV)
[2] Kousalya K, Saranya T (2021) Improved the detection and classification of breast cancer using hyper parameter tuning. Materials Today: Proceedings
[3] Shahzad A, Usman AM, Muhammad S, Anam T, Khan Shoab A (2018) Decision support system for detection of hypertensive retinopathy using arteriovenous ratio. Artif Intell Med 90:15–24
[4] Rajinikanth V, Satapathy SC, Fernandes SL, Nachiappan S (2017) Entropy based segmentation of tumor from brain mr images-a study with teaching learningbased optimization. Pattern Recogn Lett 94:87–95
[5] Dhawan AP (2011) Medical image analysis, vol 31. Wiley
[6] Houssein EH, Abohashima Z, Elhoseny M, Mohamed WM (2022) Hybrid quantumclassical convolutional neural network model for COVID-19 prediction using chest X-ray images. J Comput Des Eng 9(2):343–363
[7] Khan S, Islam N, Jan Z, Din IU, Rodrigues JJPC (2019) A novel deep learning based framework for the detection and classification of breast cancer using transfer learning. Pattern Recogn Lett 125:1–6
[8] Ucar F, Korkmaz D (2020) Covidiagnosis-net: deep bayes-squeezenet based diagnosis of the coronavirus disease 2019 (covid-19) from x-ray images. Med Hypotheses 140:109761
[9] Ucar F, Korkmaz D (2020) Covidiagnosis-net: deep bayes-squeezenet based diagnosis of the coronavirus disease 2019 (covid-19) from x-ray images. Med Hypotheses 140:109761
[10] Bergstra J, Bengio Y (2012) Random search for hyper-parameter optimization. J Mach Learn Res 13(2):281–305

[11] Ezzat D, Hassanien AE, Ella HA (2020) An optimized deep learning architecture for the diagnosis of covid-19 disease based on gravitational search optimization. Appl Soft Comput 98:106742

[12] Cuevas E, G´alvez J, Avalos O (2020) Introduction to optimization and metaheuristic methods. In: Recent metaheuristics algorithms for parameter identification. Springer, pp 1–8

[13] Hashim FA, Houssein EH, Hussain K, Mabrouk MS, Al-Atabany W (2020) A modified henry gas solubility optimization for solving motif discovery problem. Neural Comput Appl 32(14):10759–10771

[14] Hassan MH, Houssein EH, Mahdy MA, Kamel S (2021) An improved manta ray foraging optimizer for cost-effective emission dispatch problems. Eng Appl Artif Intell 100:104155

[15] Morales-Casta˜neda B, Zaldivar D, Cuevas E, Fausto F, Rodr´ıguez A (2020) A better balance in metaheuristic algorithms: Does it exist? Swarm Evolut Comput 54:100671

[16] Rojas-Morales N, Rojas M-CR, Ureta EM (2017) A survey and classification of opposition-based metaheuristics. Comput Ind Eng 110:424–435

[17] Chougrad H, Zouaki H, Alheyane O (2018) Deep convolutional neural networks for breast cancer screening. Comput Methods Prog Biomed 157:19–30

[18] http://www.wcrf.org/int/cancer-facts-figures/data- specific cancers/breast-cancerstatistics