International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 11 | Nov - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Breast Cancer Detection Using Artificial Intelligence and Machine Learning

Kanak Deepak Sahu , Ms.Namrata Soni

Department of Computer Engineering, D.Y. Patil College Of Engineering Akurdi Pune, Maharashtra, India
Department of Computer Engineering, D.Y. Patil College Of Engineering Akurdi Pune, Maharashtra, India

---***---

Abstract - Breast cancer detection using AI and ML helps inaccuracy, automates feature extraction, and allows early detection. Convolutional Neural Networks(CNNs), type of an AI model, it analyzes medical images like mammograms, ultrasounds, and MRIs to identify tumors. These helps in improving efficiency by quickly processing big datasets and assisting radiologists in making more correct diagnoses. However, data quality, bias, and ethical concerns are some challenges addressed. Overall, AI and ML offer significant potential for improving accuracy for breast cancer screening and patient outcomes.

Key Words: Artificial Decision Intelligence, Machine tree, Convutional Neural learning, Network, Early Breast Cancer Detection, Ultrasounds, Deep Learning, Accuracy, Precision, Data Privacy, Radiology, Technology.

1.INTRODUCTION

In recent research, breast cancer is one of the most common cancers. It affects thousands of women worldwide. For effective and better treatment and improvement in survival rates, early detection of breast cancer is crucial. Normally breast cancer checkup is done by radiologists or doctors manually by analyzing their medical images or reports. Mammograms, MRIs, and ultrasounds are used to find the abnormalities and lump in the breast which can cause cancer. AI and ML can assist radiologists in identifying tumors more accurately and efficiently. However, this approach for detection can be time consuming and liable to human error, or leading to missed diagnoses or false positives.

In more addition, AI-generated tools are already being used to predict breast cancer risk and suggest personalized treatments. This type of technology aims to reduce the human error, improve diagnostic accuracy, and enable early detection, so that many lives are saved. This technology represents a great chance to increase the accuracy level and efficiency of breast cancer screening.

The application of AI and ML to breast cancer identification represents a revolution in evaluation and treatment. As science technology continuously getting advance, these innovations promise to enable earlier detection, increase diagnostic accuracy, and ultimately save lives. By overcoming current barriers and encouraging collaboration between technology developers and medical the professionals, the future of cancer screening and diagnosis is bright.

This paper explores the role of AI and ML in cancer diagnosis, by focusing on their use in screening, accuracy, risk assessment, applications. It also analyzes and pathology the benefits and challenges of using this AI technology in human healthcare, as development. well as for future research and The development of computer systems that can perform tasks that require human intelligence, such as recognizing patterns and making decisions refers to AI. Machine learning (ML) is a profound extension of artificial intelligence that practices training models on big datasets. AI and ML helps to learn and study from the data and make accurate predictions without sharing specific benefits.

Additionally, limited data diversity can impact the generalizability of these models, leading to negative outcomes for unidentified populations. Algorithmic bias is another concern, as biased training data will lead to inequalities in care and equity across population groups. Additionally, the black box nature of many AI models makes it difficult for physicians to understand their decision-making processes, which can impact confidence in AI-assisted diagnosis.

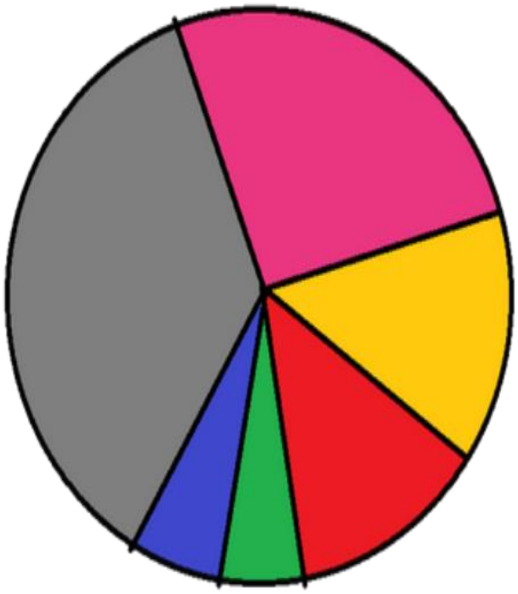

Fig -1: Cancer diagnosed chart[breast cancer second largest]

2. LITERATURE SURVEY:

1.Overview of AI and ML in Breast Cancer Diagnosis:

AI and machine learning are changing the diagnostic process by analyzing medical images. Techniques such as neural networks (CNN) are widely used for tumor detection and classification in mammography, ultrasound, machines and MRI. Research can outperform human shows that radiologists AI in some cases, especially in detecting subtle patterns in fabric text that are not visible.

2.Data Sources and Preprocessing:

The effectiveness of AI models heavily relies on the quality and quantity of data. Various datasets, such as the Digital Database for Screening Mammography(DDSM) and the Breast Cancer Wisconsin dataset, are commonly importance utilized. of data Researchers emphasize the preprocessing techniques, including normalization, augmentation, and noise reduction, to improve model performance. Several studies highlight how preprocessing can significantly enhance the accuracy of AI systems.

3.Model Development and Performance:

Many machine learning algorithms have been used to diagnose cancer, including support vector machines(SVMs), random forests, and deep learning. Many studies have shown that CNNs have the best performance in image classification due to their ability to learn hierarchical features. Many studies have shown good sensitivity and specificity, Other Cancers Breast Cancer indicating that AI models can distinguish between benign and malignant tumors.

4. Integration into Clinical Practice:
Many studies have explored the role of computer aided diagnostic (CAD) tools in medicine, introducing AI into medicine. This article demonstrates the potential of AI to reduce workload and increase diagnostic accuracy, but the also illustrates the challenges faced by physicians.

3. METHODOLOGIES USED/DISCUSSED:

The approach used to diagnose cancer through artificial intelligence (AI) and machine learning (ML)involves several key steps diagnostic accuracy and example of this process designed efficiency.

1.Databases:
Medical images are collected from a variety of sources, including hospitals, imaging laboratories, and public databases such as the Digital Screening Mammography Database(DDSM) and Wisconsin Cancer Information.

2.Data Preprocessing Normalization:
Standardize the pixel values of images to ensure consistency; this helps in learning models. Models to increase the power of the model.
Reduction: Use techniques to filter out artifacts and improve the overall image quality.
antages: Normalization ensures that gradients are consistent an elps the training process go faster. This can reduce training time, especially for gradient-based optimization algorithms.
Limitations: When features are scaled to a standard scale, the results may be more difficult to interpret. The relationship between ideas and results will become transparent.

3.Feature Extraction
Manual Feature Extraction: traditionally identified and have Radiologists extracted relevant features images. such as tumor shape and texture from(CNN) reduces the need for manual intervention by automatically learning features from raw images. Manual to extract Feature Extraction: Radiologists have traditionally identified and extracted relevant features such as tumor shape and texture from images. (CNN) reduces the need for manual intervention by automatically learning to extract features from raw images.
Working Principle: Feature extraction converts raw data into measurable features for analysis and classification in breast cancer diagnosis.
Advantages: Simple modeling and improving the performance of the business model by focusing on relevant features and automation, reducing human activity error and unfairness.
Limitations: Potential data loss, reliance on manual extraction of domain experts, and sensitivity to data changes. Additionally, automated method scan be computationally intensive and may lead to overfitting if too many features are extracted.

4.Model Development: Machine Learning Algorithm: Many algorithms are used for classification, including: Convolutional Neural Network (CNN): Very useful for photo sharing, as it can identify the spatial hierarchy in the part photo. Support Vector Machine (SVM): Can be used to classify tumors based on the extracted features. (KNN): A simple algorithm that classifies nodes based on their similarity to other examples in the database.
Working Principle: Standard development in cancer algorithm diagnosis selection, includes training data using preparation, registries, validation to and prevent interference, evaluation study before submission to the clinic.
Advantages: Model development increases accuracy, automates analysis, allows scalability for large datasets, provides instant evaluation, and enables continuous improvement by iterating with new information, leading to better performance overtime.
Limitations: Limitations of model development include reliance on good data, risk of overfitting, high cost requirements, conflicting interpretations, and bias in training information that affects the performance of the model.

4. ALGORITHMS:

Among the algorithms being used are the following:

1.Convolutional Neural Network (CNN):Description: CNNs is a special type of deep learning model. It is for specifically designed image processing. They have layers that can learn about patterns and features at different levels of abstraction. Layers include outer layers, outer layers, and full layers. The last method usually classifies the image based on the learned features. Excellence in oncology and malignancy imaging and

MRI. For example, models such as ResNet and DenseNet have been successfully used in image classification.
Limitations: High computational requirements, the need for large data sets, overexposure, lack of interpretation, and potential biases in training data suggest that these affect performance standards across individuals.

2. K Algorithm (K-Nearest Neighbors Variant)

The K algorithm* is a variation of the K-Nearest Neighbors (K-NN) algorithm and usually used for the assignment of classification problems. It works by predicting the class of some data point based on the classes of its nearest neighbors within the feature space.
Advantage : It is a non-parametric. No assumptions are made regarding any underlying distribution of the data set. Flexibility: It can fit several kinds of classification problems with the distance metric as well as by changing the number of neighbors, K.
Limitations: Sensitivity to non-uniform features, high computational cost of big data, dependence on distance measurement, sensitivity to noise in the data, and problems in efficient operation of high altitudes.
The Computational Cost: algorithm is very computationally expensive. Even though it works well even when applied to large data sets, the calculations depend on the distance between every data point.
Noise sensitivity: K* is sensitive to noisy or irrelevant features, and this can severely affect the classification accuracy.

3. Support Vector Machine (SVM)

Description: SVM is a supervised learning model that can be used for classification and regression. They work by finding the hyper plane that best separates different classes of data points in high space. stay away. This results in a robust distribution that generalizes well to unobserved data. The functions control linear and nonlinear distribution functions. They are often used in conjunction with standard subtraction techniques to classify tumors based on values obtained from image data.
Advantages: SVM is very accurate, performs well in high domain, is robust against over processing, and can handle linear and nonlinear data using kernel functions.
Limitations :SVM is considered expensive, is sensitive to outliers, requires careful parameter tuning, and can be difficult to process large datasets due to memory limitations.

4. Random Forest Algorithm:

Random Forest is a tree-based ensemble algorithm that trains several decision trees and uses their predictions for improving the accuracy of the classification and overcoming the problem of overfitting. Since its performance efficiency is improved and it can do a good amount of processing on large datasets, Random Forest is a wonderful spam.

Advantages: Accuracy: High accuracy by Random Forest mainly surpasses single decision trees or any other models.
Robustness: It has lesser overfitting tendency than a single decision tree especially if it is having a large number of trees. Manages missing data: Random Forest could easily handle missing values and was even good at maintaining accuracy if a large portion of the dataset was having missing values.
Limitations: Complexity: Despite all these powers, Random Forest can sometimes be quite computationally expensive, especially while working on big datasets and having a large number of trees.
Interpretability: Individual trees are interpretable, but the ensemble Random Forest not so much.

5. RESEARCH OUTCOMES:

The integration of artificial intelligence (AI) and machine learning (ML) in cancer diagnosis has led to many important studies that demonstrate the potential diagnosis. of the technology here to improve The application of artificial intelligence (AI) and machine learning (ML) in contributed to advances Various accuracy algorithms typically have ranges cancer screening has in accurate diagnosis. proven effective, and between 80% and95%.These diagnostic improve early diagnosis, improvements not only but also help improve patient outcomes, highlighting the important role of AI and machine learning in oncology. In some studies, convolutional neural networks (CNN) have achieved accuracy as high as 98%. Support vector machine (SVM) and random forest generally have accuracy between 85% and 90%, while the accuracy of K-nearest neighbor (KNN) is between75% and 85%.6.

6. CONCLUSION

The future of breast cancer detection through AI and ML is filled with improved patient possibility for innovation care. These technologies and can enhance early detection, personalize treatment plans, and ultimately lead to better outcomes for breast cancer patients by focusing on algorithm advancement, considerations.
Ongoing Alliance between technology developers, healthcare providers, and researchers will be essential in realizing this potential. Moreover, fostering collaboration between AI systems and medical experts can optimize decision-making processes, allowing for better diagnosis and management of breast cancer. However, to mitigate biases and promote equitable access to these innovations, considerations, ensuring addressing ethical and data security, establishing robust regulatory frameworks are

REFERENCES

[1]“Breastcancer.”https://www.who.int/newsroom/factsheets/detail/breast-cancer (accessed Dec. 03,2022).
[2] “Sanjeevani CBCC USA Cancer Hospital -Raipur.” https://scch.co.in/Cancer-Detail.aspx?I=2(accessed Dec. 03, 2022).
[3] “Statistics of Breast Cancer in India | CytecareHospitals.”https://cytecare.com/blog/breastcancer/statisticsof-breast-cancer/ (accessed Dec. 03, 2022).
[4] M. A. Richards, A. M. Westcombe, S. B. Love, P.Little johns, and A. J. Ramirez, “Influence of delay on survival in patients with breast cancer: asystematic review,” The Lancet, vol. 353, no. 9159,pp. 1119– 1126, Apr. 1999, doi: 10.1016/S0140-6736(99)02143- 1.
[5] G. N. Sharma, R. Dave, J. Sandya, P. Sharma, and K. K. Sharma, “Various types and management of breast cancer: An overview,” J Adv PharmTechnol Res, vol. 1, no. 2, p. 109, 2010, Accessed: Dec. 03, 2022. [Online].

Available:https://www.japtr.org/article.asp?issn=2231-4040;year=2010;volume=1;issue=2;spage=109;ep

6] A. Jet and H. J. O, "Supervised Machine Learning Algorithms: Classification and Comparison,"International Journal of Computer Trends and Technology, vol. 48, 2017,doi: 10.14445/22312803/IJCTT-V48P126.
[7] "Limitations of Mammograms | How Accurate AreMammograms?"https://www.cancer.org/cancer/breastcancer/screening-tests-andearlydetection/mammograms/limitationsofmammograms.html (accessed Dec. 03, 2022).