WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Machine Learning and AI-Based Detection of Breast Cancer

[1]Sohel Khan, [2]Dr.Akhil Panday, [3]Dr.Vishal Shrivastava, [4]Er.Ram Babu Buri

[1 2 3 4] Bachelor in Technology in Computer Science Engineering Arya College of Engineering and IT, Jaipur, India

ABSTRACT:

Breast cancer continues to be the most widespread and deadliest female disease globally. Detection at an early stage is a crucial step toward minimizing fatality rates and maximizing the success rate of treatments. Conventional techniques like mammography and biopsy, although useful, are mostly time-consuming and prone to human mistakes. With the arrival of artificial intelligence (AI) and machine learning (ML), the medical imaging and diagnosis sector has been revolutionized with automated, highly accurate, and scalable solutions for breast cancer detection. In this research paper, the use of different ML algorithms such as deep learning models, support vector machines (SVM), and convolutional neural networks (CNNs) for breast cancer detection is discussed. The paper emphasizes the strengths, challenges, and ways forward of AI-based breast cancer detection techniques and stresses the necessity of ethical insights, data anonymity, and explainability of the model.

Keywords: Breast Cancer, Machine Learning, Artificial Intelligence, Deep Learning, Medical Imaging, Mammogram Analysis

Introduction

Breast cancer is among the major cancer-causing causes of death for women worldwide. As per the World Health Organization (WHO), in 2020, 2.3 million women were diagnosed with breast cancer, with a total of nearly 685,000 deaths reported. Mammography and histopathological examination have been found to be effective in the early detection of breast cancer, enhancing survival. Yet, conventional methods are susceptible to diagnostic flaws in the form of false positives and false negatives. AI and ML provide novel solutions with automated diagnosis, improved accuracy, and minimized workload for radiologists.

ML-based methods use big datasets to train models that can differentiate between cancerous and benign tumors with great accuracy. In this paper, the authors debate the use of ML and AI in detecting breast cancer based on various algorithms, datasets, performance metrics, and challenges related to their usage.

Literature Review

ML and AI have received considerable interest in medical image diagnostics and analysis. Deep learning models, such as CNNs, have been shown through numerous studies to be effective in identifying breast cancer from a mammogram. For instance, Esteva et al. (2017) created an AI model that matched the classification accuracy of dermatologists in the detection of skin cancer, highlighting the potential of deep learning in medical practice.

Other studies have investigated ML algorithms like Random Forest, SVM, and Decision Trees for breast cancer classification. Litjens et al. (2017) carried out an extensive survey of deep learning applications in medical imaging, emphasizing the advantages and limitations of AI-based diagnostic tools. In spite of encouraging results, issues like data bias, interpretability, and computational demands are still major areas of concern.

Methodology

- This research entails training and testing ML models on publicly released breast cancer data like:
- Wisconsin Breast Cancer Dataset (WBCD) – Has tabular data in the form of features like tumor size, texture, and perimeter.
- Digital Database for Screening Mammography (DDSM) – Has labeled mammogram images as benign or malignant.
- Breast Cancer Histopathological Image Dataset (BreakHis) – Involves microscopic images of breast tissue samples.

Data Preprocessing

- Preprocessing is essential during ML model training to enhance accuracy and resilience. Major steps are:

- Data Cleaning: Dealing with missing values and outliers.

- Image Preprocessing: Resizing, contrast adjustment, and noise filtering.
- Feature Selection: Determining the most useful features using methods such as Principal Component Analysis (PCA).

Model Selection and Training

- A variety of ML and deep learning models were used and compared:
- Support Vector Machines (SVM) – Applied for binary classification of tumor samples. Random Forest (RF) – Ensemble learning technique for increased accuracy.
- Convolutional Neural Networks (CNNs) – Used on mammogram images to extract features and classify.
- Transfer Learning (ResNet, VGG, Inception) – Utilized pre-trained models to boost performance.

Evaluation Metrics

- The models were assessed based on:
- Accuracy – Quantifies overall accuracy.
- Precision & Recall – Evaluates reliability of the model in differentiating between benign and malignant cases.
- F1-score – Weighs precision and recall.
- Confusion Matrix – Examines classification performance.

Results and Discussion

Performance Comparison

- Experiments demonstrated that deep learning models performed better than conventional ML models in detecting breast cancer:
- CNN (ResNet-50): 95.2% Accuracy SVM: 89.7% Accuracy
- 14Random Forest: 92.1% Accuracy
- The reason for the high accuracy of CNNs is their capability to automatically learn hierarchical
- features from mammogram images. However, interpretability is an issue, requiring the incorporation of explainable AI methods.

Challenges in AI-Based Breast Cancer Detection

- Notwithstanding progress, a number of challenges remain:
- Data Unbalance: Malignant instances tend to be underrepresented, resulting in biased models.
- Model Interpretability: Deep learning models act as "black boxes," and clinicians cannot have faith in AI-driven diagnoses.
- Computational Resources: Training advanced models necessitates high GPU power, which makes them inaccessible in low-resource environments.

Future Directions

- Future studies should emphasize the following to enhance AI-based breast cancer detection:

- Designing Explainable AI: Improving model transparency with tools such as Grad-CAM and SHAP.
- Federated Learning: Overcoming privacy issues with decentralized training of AI models against datasets without any sharing of the data.
- Synergy of AI with Conventional Diagnostics: Coupling of AI with qualified radiologists towards hybrid diagnostic processes.
- Augmentation of Diversified Datasets: Making the AI models generalize for various population types to minimize bias.

Conclusion

AI and ML hold transformative promise in the detection of breast cancer, enhancing accuracy, efficiency, and accessibility considerably. Deep learning algorithms, especially CNNs, surpass conventional approaches, allowing for early diagnosis and improved patient outcomes. Model interpretability, data availability, and ethical concerns need to be overcome for extensive clinical uptake. Future development should aim at explainable AI, privacy-preserving methods, and seamless integration into clinical workflows.

REFERENCES

1. Esteva, A., Kuprel, B., Novoa, R. A., et al. (2017). "Dermatologist-level classification of skin cancer with deep neural networks." Nature.
2. Litjens, G., Kooi, T., Bejnordi, B. E., et al. (2017). "A Survey on Deep Learning in Medical Image Analysis." Medical Image Analysis.
3. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). "ImageNet Classification with Deep Convolutional Neural Networks." Advances in Neural Information Processing Systems.
4. McKinney, S. M., Sieniek, M., Godbole, V., et al. (2020). "International evaluation of an AI system for breast cancer screening." Nature.
5. LeCun, Y., Bengio, Y., & Hinton, G. (2015). "Deep Learning." Nature.