BREAST CANCER CLASSIFICATION:
A Systematic Approach Leveraging Deep Learning Models

TALLURI SASI VARDHAN
Department of CSE
Koneru Lakshmaiah Education Foundation,
Green Fields, Vaddeswaram,
A.P.-522302

THOTA BHARATH CHANDRA
Department of CSE
Koneru Lakshmaiah Education Foundation,
Green Fields, Vaddeswaram,
A.P.-522302

NAVAL DIDHARIA
Department of CSE
Koneru Lakshmaiah Education Foundation,
Green Fields, Vaddeswaram,
A.P.-522302

BAIPILLA SWAMY ESHWAR ROHANTH
Department of CSE
Koneru Lakshmaiah Education Foundation,
Green Fields, Vaddeswaram,
A.P.-522302

VEERRAJU GAMPALA

Department of CSE
Koneru Lakshmaiah Education Foundation,
Green Fields, Vaddeswaram,
A.P.-522302

Abstract

Breast cancer exhibits a Substantial threat to women worldwide, accentuating the critical need for reliable screening methods. advancements in deep learning have led to the development of efficient and accurate tools for breast cancer recognition and diagnosis. This literature review aims to comprehensively assess the current status of breast cancer detection techniques using deep learning, particularly focusing on magnetic resonance imaging (MRI). Following the guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), this review explores performance metrics, ground truths, sample sizes, MRI image types, deep learning algorithms, experimental strategies, data augmentation techniques, model architectures, training methods, and validation protocols. It analyzes the strengths and weaknesses of various approaches, identifies gaps in the literature, and offers recommendations for future research directions. The findings of this review demonstrate the significant enhancement in efficiency and accuracy of breast cancer discovery through deep learning algorithms applied to MRI data, covering advancements in preprocessing, feature extraction, and fusion techniques. Additionally, it investigates the consequence of transfer learning, ensemble learning, and multi-modal integration on diagnostic performance. This comprehensive review provides valuable insights for researchers and practitioners, enabling a enhanced comprehension of the current landscape and recent developments in deep learning-powered breast cancer diagnosis. By highlighting emerging trends, potential challenges, and areas for deeper examination , this study facilitates informed decision-making and guides future research efforts in the critical field of medical imaging.

Keywords

Machine Learning, Deep Learning, Feature Selection , Breast Cancer , Image Classification , Histopathology Image, Convolutional Neural Network , MRI , Diagnosis

Introduction

Breast cancer continues to be a significant issue on a scale, affecting a substantial number of women. Successful reduction of fatality rates associated with this prevalent form of cancer and enhancement of treatment outcomes rely greatly on the prompt and precise identification of the condition. Recent studies have delved into a variety of innovative strategies and technologies aimed at enhancing the efficiency of breast cancer detection and diagnosis.

Generative Adversarial Networks (GANs) are increasingly recognized as an effective method for overcoming the constraints posed by limited datasets when creating accurate classification models. Leveraging simulated data generated by GANs, such as the K-CGAN approach, emerges as a practical solution due to its proven ability to enhance both consistency and precision, yielding promising results in depicting the complexities found in breast cancer data. By utilizing GANs to produce synthetic samples closely resembling real data, researchers can effectively tackle concerns regarding data scarcity, enabling the development of more resilient and dependable classification models for detecting and diagnosing breast cancer. This approach not only mitigates the challenges associated with restricted datasets but also fosters advancements in machine learning methodologies for medical imaging applications [2].

Elaborate research has been carried out in the realm of-invasive methods for uncovering breast cancer early focusing on the utilization of Ultra Wideband radar systems. This novel procedure involves the use of both bistatic and monostatic radar systems in concert with advanced signal processing techniques to revolutionize tumor detection and the interpretation of reflected signals[3].

The area of medical imaging, particularly the application of advanced deep learning and machine learning techniques for the detection of breast cancer, has garnered significant attention and acclaim. Various scholarly studies have extensively delved into the categorization of breast cancer, utilizing a broad range of imaging modalities such as magnetic resonance imaging (MRI), computed tomography (CT), and sonographic imaging. The primary goal of these thorough investigations is to achieve unprecedented levels of accuracy, sensitivity, and specificity, thereby propelling the field of breast cancer analysis and treatment forward [4][5][6].

The histological diagnosis technique, widely used in detecting breast cancer, become the focus of research to improve accuracy and effectiveness. Innovative methods, such as utilitarianizing deep learning models using patches, have shown remarkable outcomes in autom the examination of histopathology images. This advancement has lessened the burden of monotonous tasks and diminished the chances of disagreements among pathologists [7].

In the realm of breast cancer prediction, progressive machine learning models like the innovative Deep Learning Assisted Efficient Adaboost Algorithm (DLA-EABA) have shown significant potential in achieving exceptionally high levels of accuracy across various imaging techniques. Additionally, the development of hybrid strategies, such as blending established algorithms like K-means and GMM (Gaussian Mixture Model), has been conducted with persistent dedication to enhancing the speed and precision of cancer detection work. These advancements underscore an ongoing commitment to enhancing diagnostic capabilities, ultimately aiming to enhance patient outcomes and advance medical research in the battle against breast cancer [8][9].

Furthermore, the incorporation of deep learning techniques, notably the transfer-learning methodology, has been studied to shorten training time and boost classification performance in breast cancer diagnosis.

Related Work

They introduced a technology known as the Knowledge-Driven Feature Learning and Integration model, or KFLI. This instrument used DWI and DCE-MRI data to identify cases of breast cancer. An adaptive weighting module and a sequence division made up the two primary parts of the KFLI. It demonstrated remarkable precision, specificity, and accuracy, making it a very flourishing tool for distinguishing breast cancer. Additionally, by emphasizing the significance of sub-sequences, it helped in the interpretation of the results [1].

The Knowledge-Driven Conditional Generative Adversarial Network, or K-CGAN, was presented as a novel model for the identification of breast cancer. Performance is enhanced by the generator loss function, which efficiently directs training by fusing binary cross-entropy and KL divergence components. Its primary goal is to improve breast cancer diagnosis; to assess its efficacy, the researchers looked at precision, accuracy, recall, and F1 Score as performance measures. Figure 11 beautifully depicts the architecture of the generator network, showcasing how this unique model incorporates both generative adversarial techniques and knowledge-driven features to effectively process information related to breast cancer. This innovative approach provides a powerful solution for enhancing breast cancer detection[2].

Our method makes use of a sharp transition bandpass FIR filter as a tool for shaping. Furthermore, we suggest that a UWB seventh derivative Gaussian pulse be used for tumor detection. Furthermore, we advise using monostatic and bistatic UWB radar equipment during the detection process. Our idea guarantees spectral mask criteria compliance with the FCC, resulting in higher spectrum use efficiency. Furthermore, our technique significantly improves the ability to reliably detect tiny cancers[3].

This research describes a revolutionary deep convolutional neural network for detecting breast cancer. The model has an AUC of 0.895, exhibiting its excellent prediction ability. With 6,132,592 trainable parameters and a training duration of about 12 hours, this network is both complicated and time-consuming to create. One distinguishing feature of the approach is the use of different models at both the breast and pixel levels, leading in improved performance. Furthermore, the heatmaps from patch categorization can be used as additional input channels, substantially enhancing the whole system [4].

They developed a new technique for detecting discriminative patches in breast histology pictures that combines a clustering algorithm with a Convolutional Neural Network (CNN) for patch selection. This approach performed admirably, obtaining an astounding 95% accuracy on the initial test set while maintaining a competitive accuracy of 88.89% throughout the test set. These findings demonstrate the efficacy of the suggested methodology, positioning it as a serious contender among cutting-edge technologies for breast cancer screening utilizing histological images [8].

Researchers have devised innovative models and tactics to enhance the detection of breast cancer in advance . One particular model called Knowledge-Driven Feature Learning and Integration (KFLI) harnesses DWI and DCE-MRI data to accomplish high levels of precision, specificity, and accuracy. Notably, the Knowledge-Driven Conditional Generative Adversarial Network (K-CGAN) distinguishes itself by integrating binary cross-entropy, leading to enhanced performance in detecting breast cancer. They have introduced a radar-centric approach that employs precise FIR filters, UWB Gaussian pulses, and bistatic/monostatic radar systems to ensure compliance with FCC guidelines and enhance the identification of small tumors. Furthermore, an advanced deep convolutional neural network demonstrates an AUC of 0.895, utilizing unique models at both breast and pixel levels to upgrade accuracy and efficiency. Finally, an integrative patch screening method combining a clustering algorithm and CNN achieves exceptional accuracy, establishing itself as a cutting-edge system for breast cancer screening with histological images. These breakthroughs collectively represent significant advancements in the field, offering a range of effective procedures for early breast cancer detection.

Datasets

The WDBC dataset is a widely acclaimed resource in machine learning and data analysis, particularly for predicting breast cancer. This dataset has 569 records and is a valuable resource for model building and evaluation. Out of these records, 212 are classified as malignant, indicating the presence of breast cancer, and 357 as benign, showing non-cancerous events. ID number, diagnosis, radius, texture, perimeter, and area are all important factors in understanding breast cancer cells. Furthermore, each image has not only a visual representation but also a collection of mean, standard error, and worst features, allowing for a more comprehensive comprehension of cancer cells' many properties. This aspect makes the dataset indispensable for future research and comparative analysis. Additionally, the material is properly organized, making it easier to research and compare breast cancer cells[14].

The Mammographic Image Analysis Society (MIAS) dataset, created exclusively for mammography research, has 322 grayscale pictures in Portable Gray Map (PGM) format. Each image is paired with a corresponding one, resulting in 161 pairs. Notably, the dataset contains precise truth data descriptions, which provide insights into irregularities and major visual aspects. This information is essential for correct analysis. Furthermore, the dataset's availability to the public stimulates widespread use and collaboration among academics. This openness fosters confidence and allows for easy comparison and replication of discoveries throughout the scientific community [12][13].

In this study, researchers utilized the BreakHis dataset, a well-known resource in medical image analysis, to categorize histological images of breast cancer. This dataset contains annotated microscopic images of breast tissue samples, which are fundamental for developing and evaluating machine learning models designed for automated breast cancer classification. The large training set of 6011 images is essential for familiarizing the model with various characteristics of breast cancer depicted in the images. The validation set, comprising 1492 images, is crucial for assessing the model's generalization to new data and pinpointing any issues like overfitting or underfitting. Lastly, the test set with 406 photos rigorously evaluates the model's performance and adaptability in real-world scenarios, providing an unbiased assessment of its effectiveness on unseen data. [7][4].

Methodology

The literature review focused on utilizing deep learning and texture analysis to detect breast cancer from MRI images by excluding articles unrelated to these topics. It examined various deep learning algorithms, analysis methods, MRI image variations, ground truth data, sample sizes, and lesion quantities. The review offered valuable insights, identified challenges, and proposed potential research directions. It emphasized the importance of deep learning in enhancing the faithfulness of breast cancer diagnosis through MRI images to improve patient outcomes. By systematically reviewing literature from diverse databases following PRISMA guidelines, the study revealed both advancements and limitations in using deep learning for breast cancer detection. Furthermore, it highlighted the crucial role of machine learning in aiding radiologists to interpret breast MRI images accurately and efficiently, with the technology's potential to raise diagnosis authenticity and streamline clinical workflows [1].

The revolutionary K-CGAN method was utilized to create synthetic data for classifying breast cancer. Through a detailed assessment of essential performance measures like recall, precision, accuracy, and F1 Score, the research substantiated the exceptional efficacy of K-CGAN over other existing GAN versions. The results highlighted the effectiveness of K-CGAN in optimizing breast cancer classification by generating artificial data. This technique not only surpassed alternative GAN variations but also demonstrated its capability to strengthen the precision and dependability of breast cancer detection models [2].

In the context of breast cancer detection, the average accuracy scores for the classifiers are as follows:

Decision Tree: 92.4%

Random Forest: 96.5%

Support Vector Machine: 93.6%

These scores indicate the overall accuracy of each model in correctly predicting the presence or absence of breast cancer in unseen data.

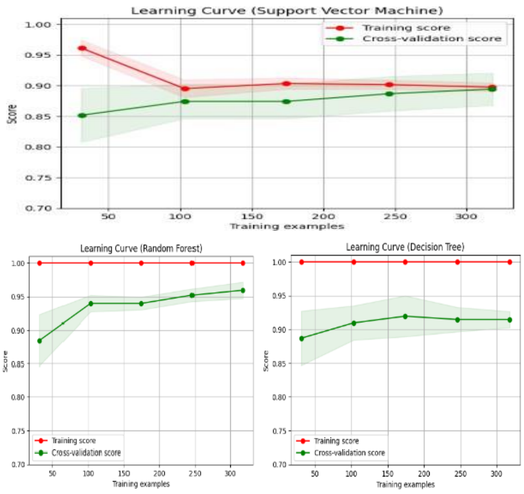

Figure:Average Accuracy scores on various models.

They created a deep convolutional neural network to classify breast cancer screening exam findings.

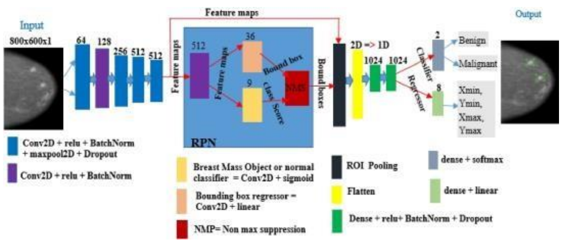

Figure:Structure of Breast Cancer Detection architecture flow work[15]

The utilization of the Adam optimization method, L2 regularization and specific models at both breast and pixel-levels led to significant achievements in the accurate prediction of cancer presence. Furthermore, a deep convolutional neural network was developed using the WDBC dataset, accompanied by a visual representation demonstrating the dataset's precision. [5].

The deep convolutional neural network was developed using the WDBC dataset, with a graphical representation illustrating the dataset's accuracy. The model achieved a significant accuracy score, outperforming other CNN models like WOA-CNN, GA-CNN, PSO-CNN, and GWO-CNN. The study highlighted the importance of deep learning algorithms in enhancing breast cancer detection accuracy and emphasized the significance of model optimization for improved performance.

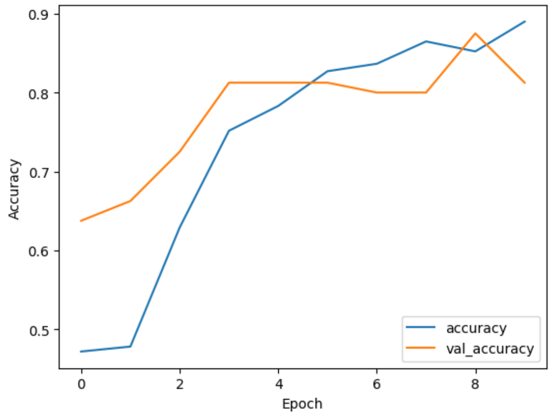

Figure:Process for deep convolution neural network

Authors Name	Methodology	DataSet	Accuracy
Jing Zheng, Denan Lin, Zhongj Gao, Shuang Wang, Mingjie He, Jipeng Fan	DLA-EABA algorithm achieved high accuracy of 97.2%. Used CNNs for tumour classification.	MIAS data set	97.2%
Huan-Jung Chiu, Tzuu-	Proposed method achieved an	Dataset from Manuel Gomes,	**86.97 %**

Hseng Li, Ping-Huan Kuo	accuracy of 86.97%. Used PCA, multilayer perceptron, and transfer learning with SVM.	University Hospital Centre of Coimbra	
Emilija Strelcenia, Simant Prakoonwit	K-CGAN method for synthetic data generation. Evaluation with five classification methods and feature selection.	Wisconsin Breast Cancer Dataset	High accuracy
Yasin Yari, Thuy Nguyen, Hieu Nguyen	Achieved up to 98% accuracy in multiclass classification. Employed bottom-up development approach and Agile SCRUM methodology.	BreakHis dataset	Up to 98%
Varsha Nemade Vishal Fegade	Decision tree and XGBoost classifier have the highest accuracy of 97%.	Wisconsin Breast Cancer Dataset	highest AUC-ROC of 0.999

LIMITATIONS

One significant obstacle faced in the realm of medical artificial intelligence is the scarcity of extensive and varied datasets that are meticulously annotated. This creates notable challenges in attaining precise annotations for medical images. Consequently, numerous models are trained on datasets lacking in diversity, resulting in biased results. Moreover, the process of comprehending and elucidating the decision-making mechanisms of deep learning models is complex.

Additionally, there is a restriction on the suitability of these models to new datasets and patient characteristics. These issues raise ethical concerns about the safety, accountability, privacy, impartiality, and transparency of deep learning models in the healthcare industry [1].

The scarcity of training data presents a significant hurdle to the advancement of deep learning in medical imaging. Constructing extensive datasets for medical imaging poses a formidable challenge, requiring considerable time and resources. Moreover, effectively evaluating the efficacy of deep learning models across various research pursuits remains a complex task [4].

The scarcity of medical data impedes the competence of deep learning analysis. This challenge is exacerbated in Africa by the shortage of pathologists. Furthermore, the risk of human errors presents an additional obstacle to achieving accurate diagnoses. Moreover, the BreakHis dataset, employed to train a network from the ground up, is relatively small. Adding to the complexity, the proposed models have exhibited an anomaly in the form of false-negative predictions [7].

The Proficiency of the model may be impacted by the restricted availability of hardware resources such as GPU. Presently, the model is constrained to the classification of cancer regions in a binary manner. However, there exists potential for the model to be expanded upon with the aim of classify various types of cancers[10].

CONCLUSION

Recent advances in AI, machine learning, and deep learning have opened up a new field in breast cancer research. Key findings include promising results from GANs, notably K-CGAN, for breast cancer classification, as well as the possibility for synthetic data to improve model training. Additionally, ultra-wideband radar appears as a novel, non-invasive method for early detection, with a focus on improving tumor location accuracy.

Various machine learning and deep learning approaches, such as SVM, DT, Neural Networks, CNNs, and ensemble methods, offer diverse tools for breast cancer detection. The development of sophisticated AI systems holds significant potential for prediction and prevention.

In radiology, deep neural networks show proficiency in classifying screening exams, and hybrid models combining radiologists' expertise with neural networks enhance diagnostic performance. Patch-based models, like Pa-DBN-BC, outperform whole-image models, emphasizing the importance of localized feature extraction.

The use of hybrid methodologies, such as combining K-means and GMM with transfer learning, has been proven to improve precision in breast cancer diagnosis, as demonstrated by the successful implementation of the VGG16 model. However, problems remain in the ethical, legal, and social spheres, necessitating careful consideration before widespread clinical implementation. Future research should combine clinical data, measure health consequences, and solve dataset restrictions in order to make significant advances in the field.

The inclusion of many evaluation metrics, such as accuracy, sensitivity, specificity, AUC, and F-score, emphasizes the multifaceted character of breast cancer detection research. The integration of AI technologies presents a vast potential for enhancing breast cancer treatment. This convergence of technologies offers a promising avenue for improving the accuracy and capability of breast cancer diagnosis and personalized medicine. Collaborative efforts and novel approaches contribute to continuous progress, but as these technologies move toward broader clinical application, careful navigation of difficulties and ethical issues is critical.

REFERENCES :

[1] Richard Ada, Kevin Dell'Aquila, Laura Hodges1, Takouhie Maldjianand Tim Q. Duong ,2023. Adam et al. Breast Cancer Research

[2] Emilija Strelcenia ,Simant Ptrakoonwit ,2023.Improving Cancer Detection Classification Performance Using GANs in Breast Cancer Data , Department of Creative Technology, Bournemouth University, BH12 5BB Poole, U.K, IEEE.

[3] M.K. Devika Menon and Joseph Rodrigues ,Efficient Ultra Wideband Radar Based Non Invasive Early Breast Cancer Detection,2023. Department of Electrical and Electronics Engineering, CHRIST (Deemed to be University), Bengaluru 560074, India.

[4] ReemJalloul , H. K. Chethan and Ramez Alkhatib, A Review of Machine Learning Techniques for the Classification and Detection of Breast Cancer from Medical Images.,2023.

[5] Nan Wu Thibault Févry, Joe Katsnelson, Eric Kim , Jungkyu Park, Yiqiu Shen , Zhe Huang, Masha Zorin, Stanisław Jastrze˛bski, , Stacey Wolfson, Ujas Parikh, Sushma Gaddam, Leng Leng Young Lin, Kara Ho, Joshua D. Weinstein, Beatriu Reig, Yiming Gao, Hildegard Toth, Kristine Pysarenko, Alana Lewin, Jiyon Lee, Krystal Airola, Eralda Mema, Stephanie Chung, Esther Hwang, Naziya Samreen, S. Gene Kim, Laura Heacock , Linda Moy, Kyunghyun Cho, and Krzysztof J. Geras,Deep Neural Networks Improve Radiologists Performance in Breast Cancer Screening, IEEE TRANSACTIONSON MEDICAL IMAGING, VOL. 39, NO. 4, APRIL 2020

[6] JING ZHENG MINGJIE HE,DENANLIN ,ZHONGJUN GAO,ANDJIPENG FAN ,SHUANGWANG ,Deep Learning Assisted Efficient AdaBoost Algorithm for Breast Cancer Detection and Early Diagnosis,IEEE ,May 8,2020.

[7] IRUM HIRRA, MUBASHIR AHMAD ,IFTIKHAR AHMED SAEED ,AYAZ HUSSAIN, M. USMAN ASHRAF ,SYED FURQAN QADRI , AHMED M. ALGHAMDI ,ANDAHMEDS.ALFAKEEH , DeepLearning Applied for Histological Diagnosis of Breast Cancer, IEEE, February 2, 2021

[8] YUQIAN LI ,JUNMIN WU AND QISONG WU, Classification of Breast Cancer Histology Images Using Multi-Size and Discriminative Patches Based on Deep Learning, February 7, 2019,IEEE.

[9] HUAN-JUNG CHIU ,TZUU-HSENG S. LI ,AND PING-HUAN KUO,Breast Cancer Detection System Using PCA Multilayer Perceptron Transfer Learning and Support Vector Machine, November 10, 2020,IEEE.

[10] USMAN NASEEM,QAZI EMAD UL HAQ ,JUNAID RASHID ,LIAQAT ALI, JUNGEUN KIM 5,MAZHARJAVED AWAN,AND MUHAMMAD IMRAN,Breast Cancer Classification From Histopathological Images Using Patch-Based Deep Learning Modeling,IEEE, 12 May 2022.

[11]-An Automatic Detection of Breast Cancer Diagnosis and Prognosis Based on Machine Learning Using Ensemble of Classifiers

[12] P. ESTHER JEBARANI, N. UMADEVI1, HIEN DANG AND MARC POMPLUN,A Novel Hybrid K-Means and GMM Machine Model for Breast Cancer Detection,October 27,2021.IEEE

[13] ABEER SABER ,MOHAMEDSAKR, OSAMA M. ABO-SEIDA, ARABI KESHK, AND HUILING CHEN ,A Novel Deep Learning Model for Automatic Detection and Classification of Breast Cancer Using the Transfer-Learning Technique,IEEE, May 11, 2021.

[14] Varsha Nemade1,Vishal Fegade, Department of Mechanical Engineering, Mukesh Patel School of Technology Management and Engineering ,NMIMS Shirpur Campus, India. ELSEVIERM ,PROCEDIA,2023.

[15] Simon Hadush Nrea, Yaecob Girmay Gezahegn, Gebrekirstos Hagos, Abiot Sinamo Boltena, BREAST CANCER DETECTION USING CONVOLUTIONAL NEURAL NETWORKS. ICLR ,2020.