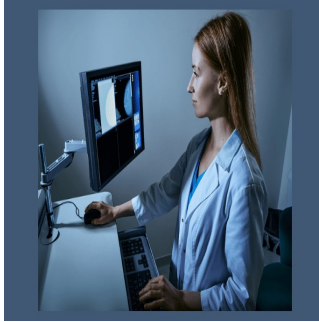## AI in Breast Cancer Screening: Progress, Challenges & Future Directions

The automated detection of breast cancer in mammography using computer algorithms has been a research focus for decades, aimed at assisting radiologists in identifying early signs of cancer. Despite advancements, breast cancer screening using digital mammography (DM) or digital breast tomosynthesis (DBT) still heavily relies on radiologists, who can make errors such as false negatives and false positives. Screening programs are also labour-intensive and expensive, often requiring multiple radiologist reviews, particularly in Europe. To improve accuracy and reduce workload, researchers have explored computer-aided detection (CAD) systems since the 1990s. Early CAD systems using traditional machine learning were not widely accepted due to high false positive rates. Deep learning (DL) AI techniques have recently emerged, showing significant improvements in detection performance and gaining praise for potentially matching or surpassing radiologists' accuracy. However, further research is necessary for their successful integration into clinical practice. This review article published in the European Journal of Radiology discusses the current state of AI technology in breast cancer detection, available systems, and the ongoing challenges in the field.

Evolution and Impact of Deep Learning in Breast Imaging

AI has significantly advanced in medical imaging, particularly breast imaging, due to affordable computing power, digital data, and cloud resources. AI encompasses a range of techniques designed to simulate human intelligence, including machine learning (ML) and deep learning (DL).

Traditional computer-aided detection (CAD) systems, which emerged in the 1990s, used traditional ML strategies. These systems relied on manually extracted image features that were processed through models like decision trees and support vector machines. Such approaches required large amounts of structured training data and human supervision, as most learning was done in a supervised manner.

In contrast, DL-based CAD tools, which belong to a more recent subtype of ML, use neural network architectures to create models capable of making accurate data-driven decisions. Unlike traditional ML, DL does not require manual feature selection; instead, features are learned and optimised during training through backpropagation. DL includes various forms such as deep neural networks (DNN), recurrent neural networks (RNN), deep belief networks (DBN), and convolutional neural networks (CNN). CNNs, in particular, have proven highly effective for image segmentation and classification, with newer architectures like Transformers significantly improving upon CNN performance.

The ability of DL models to learn relevant features directly from datasets has revolutionised medical image analysis. Neural networks have achieved detection rates comparable to average radiologists. Furthermore, DL has been shown to reduce radiologist workload and improve cancer detection sensitivity without loss of specificity in randomised screening trials. DL also enables the development of classification models that distinguish between malignant and benign lesions, reducing variability among observers.

DL-based AI systems are widely used in breast cancer detection across various imaging technologies such as digital mammography (DM), digital breast tomosynthesis (DBT), and breast MRI. As medical imaging becomes increasingly complex with 3D and 4D technologies, AI will play a crucial role in processing large datasets. The current trend in AI is to combine multi-modal input data with non-imaging data to enhance accuracy and outcomes.

The Vital Role of Data Pre-processing in AI Development

Data is often compared to oil in the 21st century due to its immense value. However, like oil, raw data needs extensive processing to become useful. Data scientists spend up to 60% of their time on data cleaning and processing in data mining projects. The quality of data is crucial for AI algorithms, as they rely on this data to identify patterns and features for accurate outcomes. Pre-processing strategies, including data acquisition, curation, annotation, and storage while maintaining patient privacy, are essential for the performance of AI algorithms, especially in

medical imaging. Combining imaging and non-imaging data enhances the clinical relevance of AI models, necessitating meticulous attention during pre-processing. Despite this, the success of AI models also hinges on robust model development and training, as sophisticated models cannot compensate for poor data quality.

AI algorithms use several learning paradigms: supervised, unsupervised, semi-supervised, and reinforcement learning. In supervised learning, data is fully annotated with known ground truths, such as labelling mammograms with "cancer" or "no cancer" for detection tasks or providing bounding boxes for lesion segmentation. Unsupervised learning models use untagged images to find common patterns, while semi-supervised learning combines elements of both supervised and unsupervised learning, utilising both labelled and unlabelled data. Reinforcement learning involves an agent learning through trial and error in an interactive environment, using feedback from its actions to improve.

AI development typically involves splitting data into training, validation, and test datasets. When datasets are insufficiently large, data augmentation strategies, such as rotation and scaling, are employed to increase sample sizes without affecting the algorithm's expected outcomes. Synthetic data generated by adversarial networks can also augment data, especially for rare cancer types or less frequent classes. Transfer learning, where pre-trained networks are fine-tuned with new but limited data, is another common strategy, requiring less training time for effective performance in a new domain.

Enhancing Trust in AI for Healthcare: Guiding Principles and Initiatives

AI tools, especially those based on deep learning (DL), are often perceived as black boxes, with results that seem to be generated magically, which reduces trust and hinders clinical implementation. This scepticism is prevalent despite the excitement surrounding AI in healthcare, as many CE-marked AI products lack scientifically proven evidence of their clinical efficacy. To enhance trust in AI, the European Commission published the Assessment List for Trustworthy AI (ALTAI) in 2020, though these guidelines are not specific to healthcare. Consequently, several international initiatives have proposed checklists and guidelines tailored to medical imaging, such as TRIPOD-AI/PROBAST-AI, CLAIM, MINIMAR, CONSORT-AI, CLEAR, Metrics Reloaded, and FUTURE-AI. The FUTURE-AI guidelines, for example, emphasise six principles to improve clinical trust: Fairness, Universality, Traceability, Usability, Robustness, and Explainability.

Fairness involves ensuring consistent AI performance across individuals and subgroups, mitigating biases through strategies like data augmentation for under-represented groups. Universality promotes the application of standards during AI development, evaluation, and deployment to enhance interoperability and applicability across clinical settings. Traceability requires mechanisms for documenting and monitoring AI development and performance in clinical environments, allowing for maintenance interventions. Usability ensures that AI tools are practical and acceptable for real-world users, involving evaluations that include all end users. Robustness maintains AI performance under variable real-world conditions, necessitating the use of multi-centre and multi-vendor datasets during development and validation. Explainability provides meaningful insights into AI predictions, enhancing user trust by elucidating the decision-making processes.

Developing independent and generalisable AI models is crucial for clinical implementation. This requires training with high-quality, unbiased data, often from multicenter and multivendor datasets, to ensure generalizability. Privacy issues, however, can complicate access to comprehensive datasets. Techniques like federated or swarm learning address these challenges by allowing AI models to be trained locally without data sharing, combining models from different centres without transferring sensitive information.

AI-powered results must be interpretable for both developers and clinical users. Explainable AI (XAI) aims to make AI predictions understandable, improving trust and aiding in clinical decision-making. Interpretability models, such as Grad-CAM, use visual explanations like heat maps to highlight important image features considered by the AI model, making the decision-making process transparent.

Implementing AI in Breast Cancer Screening: Challenges and Opportunities

AI systems can be implemented in various ways in breast cancer screening using digital mammography (DM) or digital breast tomosynthesis (DBT), depending on local needs and preferences of each screening program. The potential impact of AI on screening varies with its specific implementation. Some AI approaches can significantly reduce screening workload without compromising sensitivity, while others may increase sensitivity but also raise false positive rates. The ideal improvement in workload, sensitivity, and specificity through AI in screening is still being validated, although some studies, like those by Lång et al., show promising results.

These AI applications are not mutually exclusive and can be combined. Despite extensive research in breast imaging AI, evidence remains limited. Most studies are retrospective, involve only a few commercially available AI systems, and use limited data from single sites, lacking the heterogeneity necessary to represent global screening programs. Other common limitations include dataset enrichment, unconsidered changes in radiologist behaviour with AI involvement, and limited clinical relevance of results, such as the aggressiveness of screen-detected cancers.

Each system's characteristics, including its regulatory-approved intended use and performance features, also constrain the use of breast imaging AI. For instance, an AI system cleared only for concurrent support during mammogram interpretation cannot be used as a stand-alone reader in screening, even in a double-reading scenario.

Navigating AI Adoption in Breast Cancer Screening

♦ Choosing the Right AI System for Breast Imaging

With over half a dozen breast imaging AI systems available, users face the challenge of selecting the most suitable system for their clinical setting. While some systems are more accurate than others, factors such as intended use, deployment possibilities, and specific performance for the user's population and equipment are crucial. Periodic evaluations or audits of AI algorithms are recommended to ensure safety and effectiveness. AI developers are encouraged to provide tools for performance monitoring to enhance transparency and aid in decision-making. Well-established guidelines and checklists will be essential for improving the trustworthiness of future AI algorithms, with traceability measures in place for documenting and monitoring AI tools in clinical environments. Internal validation using local data is crucial to detect potential biases.

♦ Technological Improvements for AI in Breast Cancer Screening

Current AI algorithms for breast cancer detection typically use a single DM or DBT image, but radiologists consider multiple data sources, such as prior images and patient records. Future AI algorithms should integrate multi-source data to make more accurate predictions. This requires access to more clinical data and strategies like federated learning or synthetic data generation to preserve patient privacy.

♦ Staying Updated with AI Developments

The field of AI is rapidly evolving, with extensive information available in radiology journals and conferences. However, there is a gap between controlled experimental results and real-life clinical use. Aggregating and publicising performance benchmarks of AI in clinical settings, similar to the Breast Cancer Surveillance Consortium, can provide valuable insights into the actual impact of these algorithms.

♦ Prospective Clinical Trials and Real-Life Evidence

Most evidence on the impact of AI in breast cancer screening is based on retrospective data or laboratory studies. Prospective clinical trials are essential for widespread adoption. Recent trials like MASAI and ScreenTrustCAD have shown promising results, validating retrospective findings. The ongoing AITIC trial aims to evaluate the use of AI in real-life screening, suggesting that AI can safely and effectively label low-risk exams as normal, increasing cancer detection rates without decreasing recall positive predictive value. The question remains whether regions will leverage evidence from other trials or conduct their own, potentially delaying AI benefits. Nonetheless, many hospitals are already implementing AI based on local clinical needs.

Paving the Way Towards Breast Cancer Detection Revolution

The use of AI technology for automated breast cancer detection in mammography has made significant strides and offers great potential for enhancing screening programmes. Traditional CAD systems employing machine learning have been around for decades, but their high false positive rates limited their acceptance. Recent advancements in deep learning (DL)-based AI have shown improved performance, sometimes surpassing radiologists in breast cancer detection.

DL-based AI systems in breast cancer screening using digital mammography (DM) or digital breast tomosynthesis (DBT) have been extensively researched. These systems aid radiologists by improving accuracy and reducing false negatives and false positives. There is even potential for AI to replace radiologists in reading mammograms, although more research is necessary to ensure successful implementation.

Training AI systems requires high-quality data, with significant emphasis on data pre-processing for optimal performance. AI utilises various learning paradigms, including supervised, unsupervised, semi-supervised, and reinforcement learning, along with data augmentation and transfer learning to address data limitations.

Validation is critical to ensure the trustworthiness and clinical efficacy of AI systems. International guidelines and checklists are being developed to improve transparency and reliability, which will foster greater acceptance and integration of AI tools in clinical practice. Despite the promise of AI in improving breast cancer detection, ongoing research, collaboration, and validation are essential for its successful integration into routine screening programmes.

Source: [European Journal of Radiology](#)

Image Credit: [iStock](#)

Published on : Thu, 23 May 2024