https://africanjournalofbiomedicalresearch.com/index.php/AJBRr
Afr. J. Biomed. Res. Vol. 27 (September 2024); 693-699
Research Article

Integration of Artificial Intelligence in the Diagnosis of Breast Cancer Using 3D Mammography

Anil Kumar[1*], Dr. Keya De Mukhopadhyay[2], Dr Ashok Kumar Verma[3], Dr Pushpa Mamoria[4], Dr Wali Tariq[5], Mr. Dawn V J[6]

[1]Assistant Professor, Department of Anatomy and Neurobiology, College of Medicine and Health Science (CoMHS), National University of Science and Technology (NUST), Sohar, Al Tareef. Postal code: 391, P. O. Box: 321, Sultanate of Oman, Sultanate of Oman, E-mail: anilkumar@nu.edu.om*
[2]Associate Professor, Department of Biotechnology, Institute of Engineering and Management, New Town, University of Engineering and Management, Kolkata Keya.DeMukhopadhyay@uem.edu.in
[3]General and Laparoscopic Surgeon national board of examination (NBE), ashokasdoc@yahoo.com
[4]Department of Computer Application, School of Engineering and Technology, Chhatrapati Shahu Ji Maharaj University, Kanpur, p.mat76@gmail.com
[5]Dean Academics, Associate Professor, KM Medical College & Hospital, MATHURA-281 123. drstwali@gmail.com
[6]Associate Professor, Department of Pharmacy Practice, Sanjo College of Pharmaceutical Studies Palakkad,Kerala, 678702 dawnvjvj@gmail.com

Abstract
Breast cancer remains one of the leading causes of cancer-related deaths among women globally. Early detection through advanced imaging techniques is crucial for improving survival rates. This study evaluates the integration of an artificial intelligence (AI) system into 3D mammography (digital breast tomosynthesis, DBT) for breast cancer diagnosis. A total of 5,000 women who underwent DBT screening between January 2021 and June 2023 were included. The AI system, based on a deep learning convolutional neural network, was trained on a dataset of annotated DBT images and compared against the interpretations of two experienced radiologists. Key performance metrics such as sensitivity, specificity, accuracy, and area under the curve (AUC) were analyzed. Results showed that the AI system achieved higher sensitivity (94.2%) and specificity (92.5%) than the radiologists, with an AUC of 0.968, indicating superior diagnostic performance. Additionally, AI-assisted readings significantly reduced radiologist interpretation time by 44%, suggesting workflow efficiency improvements. While the AI system showed promising results in improving detection accuracy and efficiency, further studies in diverse populations are needed to validate its clinical application. This research highlights the potential of AI as a valuable tool in breast cancer diagnosis, aiding radiologists in enhancing diagnostic accuracy and reducing time to diagnosis.

Keywords: artificial intelligence, breast cancer, 3D mammography, diagnostic accuracy, deep learning

**Author for correspondence: anilkumar@nu.edu.om*

Received: 23.09.2024 Accepted: 10.10.2024

DOI: https://doi.org/10.53555/AJBR.v27i3.2700

Introduction

Breast cancer to date is considered one of the most prevalent types of cancer in women all over the world. The data provided by the World Health Organization (WHO) reveal that in 2020, there were more than 2. 3 million new cases of breast cancer making the disease the most common cancer all over the world. Mammography has been the key modality for breast cancer screening, and early detection is a key determinant in improving survival rates. Nevertheless, similar to any other diagnostic tools, the sensitivity and specificity of mammography are not perfect, especially in women who have dense breast tissues. Screening using conventional two-dimensional mammography causes overdiagnosis or overlooking of cancer leading to delayed diagnosis. Three-dimensional mammography or digital breast tomosynthesis has increased the breast cancer detection rate, especially in women with dense breasts [1]. The introduction of AI into this diagnostic modality brings light to the possibility of improving the detection of breast cancer and minimizing errors associated with human activity. It took decades to go from transforming the traditional x-ray mammography to using AI in combination with the latest technology i.e., 3D mammography in diagnosing breast cancer. AI in the form of machine learning and deep learning algorithms can give detailed assessment of imaging data sets with high precision with less bias as seen in human decision-making. This integration enhances the diagnostic capability besides having the benefit of increasing the current diminishing supply of radiologists globally. Breast cancer's prevalence is still on the rise, sometimes even by deadly methods; as such, the application of technologies like AI in 3D mammography could be one of the solutions for better diagnosis [2].

The Global Burden of Breast Cancer

Breast cancer currently ranks as one of the major causes of cancer-related deaths among women in the world. Breast cancer: Statistics, risks and types Breast cancer is the second most common cancer among women in the United States with a lifetime risk of 1 in 8 to develop the disease, according to estimates by the American Cancer Society. Ageing populations, changes in lifestyles and increased awareness of women's genetic risks are only some of the causes as to why breast cancer frequency is on the rise. Screening programs have been advanced to have the outcome of reducing mortality since they enable treatment of the condition at early stages. Mammography, which was developed in the mid-twentieth century has continued to be the best method of breast cancer screening [3]. However, it is not without limitations especially when it is used in the conventional Two Dimensional form. High density where the breast tissue appears white masks tumours that appear white in the mammograms thereby making it difficult to identify malignancies.

Evolution from 2D Mammography to 3D Mammography

- Two-dimensional digital mammogram as the name suggests relies on producing mammographic images that are planar, displaying all the tissue structures in a single plane. Though this method has been proven viable, it presents a lot of difficulties, particularly for women with dense breast tissue mass; this makes the image appear similar because the tissues obscure one another. This often leads to what is known as 'overdiagnosis' where many patients are subjected to biopsies or extra imaging, only to find that they do not have cancer or 'over-treatment', where patients are treated for cancer when in fact they do not have it. Overcoming these shortcomings 3D mammography was invented [4].
- Breast tomosynthesis or three-dimensional mammography acquires different pictures of the breast creating thin slices of the breast tissue. This technique helps in the soeparation of various structures in the breast, this minimizes the overlying tissue which may hide cancers. According to Skaane et al. [4], 3D mammography improves cancer detection by 27% than 2D mammography. In addition, it decreases recall rates such that they can be lowered by up to 15%, which translates to lower false positive rates. However, as indicated in the preceding sections, DBT is not immune to some challenges, which include; the need for specific equipment, high radiation doses and, time spent by radiologists in analyzing images.

The Role of Artificial Intelligence in Medical Imaging

- In medical imaging the application of AI especially machine learning and deep learning has been shown to have a lot of potential. A different approach is machine learning which uses algorithms that can recognize patterns in the data set and will perform better with every repetition without being programmed. ML has several classifications: Deep Learning uses an artificial neural network that is designed to mimic the operation of the human brain to analyse patterns. Such technologies are very useful in various applications including image classification, segmentation and anomaly detection among others.
- In breast cancer, screening requires the use of AI to train software to read mammographic images distinguish between possible malignant lesions and characterize the lesions according to the probability that they are malignant. Due to this fact, AI algorithms are capable of analyzing thousands of images in just a short period compared to human radiologists this makes it the best tool to use when carrying out large-scale screening. Another benefit of AI is the learning ability of the system where it will be able to extract information from a large database and come up with patterns that a human eye may not be able to detect. A review of the study by Rodriguez-Ruiz and his colleagues [4] indicates that AI was equally effective or even superior to experienced radiologists when it came to identifying breast cancer through mammography, especially in cases involving dense breast tissue.

AI Integration in 3D Mammography: A Transformative Approach

There is a significant capability for AI to advance the present 3D mammography lecture by giving an increase in exactness, speed and availability. AI can assist radiologists in several ways, including

1. Automated Detection and Diagnosis: Computer-assisted detection in 3D mammography can be employed to teach algorithms to identify patterns related to malignancy, to

highlight certain areas that may be suspicious. This, in turn, lessens the burden on the radiologists and helps in avoiding errors that may be prone to humans. McKinney et al. similarly showed that the use of artificial intelligence could lower the false positive rate by 5. 7% and false negatives by 9 % [5]. In another clinical study aimed at comparing the effectiveness of the normal NP testing method and the new NP testing method proposed by the authors, the results achieved show that the sensitivity of the normal NP testing method in detecting positive cases in this study is lower by 7% and false negatives are lower by 9. 4 per cent, which upon comparison to the previous method indicates a 4 per cent increase in efficiency of diagnosis.

2. Reduction of Radiologist Workload: There is a need to screen breast cancer especially because the disease is common and the workforce of professional radiology technicians is limited and could be overwhelmed thereby developing cases of Burnout. They found that for lung nodule detection, AI can help as a 'second reader' and can pinpoint the areas that need special review from radiologists. This makes the job of radiologists easier and easier to concentrate on the difficult areas which an AI can easily analyze. Lehman CD et al. estimated that an AI-based approach to reading mammography could reduce workflow time by approximately 44% [6].
3. Personalized Risk Assessment: AI is also capable of offering risk scores based on the results of the patient's mammogram, or with other factors like age, family history, and breast density. When combined with mammographic images, the aforementioned factors will enable the application of AI to distinguish women with altered risk probability of breast cancer and modify the screening regularity. Yala et al. noted that the adoption of the AI risk models can enhance the performance of traditional risk assessment hence promoting early screening and intervention [7].
4. Improved Detection in Dense Breast Tissue: It is traditionally known that it is difficult to facilitate early breast cancer diagnosis in women with dense breast tissue since such tissue can obscure tumours on a mammogram. The use of AI has demonstrated a high degree of accuracy in detecting cancer in women with dense breast tissue since it can detect even slight changes in the pattern of the tissue. Conant et al. (9) found that there was an increase in cancer detection by 30 % when using AI-assisted 3D mammography for women with dense breasts as compared to 2D mammography.

Challenges and Ethical Considerations in AI Integration

- Even though using AI in the diagnosis of breast cancer has several benefits as mentioned above, there are several challenges and ethical issues that need to be met for the application of this technology to be common. Another one is the lack of transparency that is characteristic of many AI algorithms and programs; the internal mechanisms of decision-making are often opaque or hidden. However, clinical natural language processing lacks interpretability and thus calls into question, a certain level of accountability every time there is a diagnostic mistake. Also, the machine learning algorithms are subservient to the training set, in that, if the training set has some form of prejudice will be duplicated in the end product. For instance, it is illustrated that AI models trained with the data belonging to Caucasian women can be less accurate about women of other ethnicity [7b].
- Some of them include; Overconfidence; on this, the use of AI could make radiologists lazy and rely entirely on the output generated by the AI. AI can help in making a diagnosis but it cannot act as a part of a radiologist's comprehensive experience. It is important to guard against situations where artificial intelligence replaces human expertise which is very crucial in the provision of quality care.

The Future of AI in Breast Cancer Diagnosis

Down the line, the advancements in the field of AI are expected to be seamlessly integrated into breast cancer diagnosis. These algorithms are still under development, and there are many opportunities for improving machine learning algorithms to allow for the connection of multiple data sources, including genomic data and patients' medical records to provide better and more individualized cancer treatment. In addition, the growth of Explainable AI that utilizes methods that make it clear to identify why the given conclusion has been made can contribute to concerns about transparency and accountability. It will also help to improve accessibility and availability of good quality breast cancer screening especially in developing countries where there are few radiologists. Advanced deep learning applications might let healthcare providers in telemedicine-inefficient or low-population density zones perform breast cancer screening without having a full-fledged team of radiologists [8].

Materials and Methods

Study Design

In this study, the performance of an AI model in breast cancer diagnosis based on 3D mammography the digital breast tomosynthesis (DBT) was assessed using a retrospective review design. The objective was to compare the performance of using an AI system to that of the standard radiologist interpretation of accuracy, sensitivity, specificity, as well as diagnostic efficiency. In more detail, before data collection, consent to participate in the study was obtained from the Human Research Ethics Committee (IRB) and the guidelines concerning the treatment of patient data were observed.

Study Population

The study involved an estimated sample of 5000 premenopausal and postmenopausal women who had 3D mammography screenings in a tertiary care hospital over the 18 months of January 2021 and June 2023. Participants were selected based on the following inclusion criteria:

1. Female clients that have had at least one DBT screen in the said period.
2. Full clinical information; repeat result of biopsy results.
3. Does not have any history of breast cancer or any kind of surgery related to breast cancer.

Exclusion criteria included:

1. Those patients who didn't do their imaging or have their medical history fully documented.

2. Pregnant women, breastfeeding mothers, and women with prior breast operations or biopsy that will skew the images.
3. Women with other diseases like malignancy in other sites, or breast diseases not originating from cancer.

Imaging Data Collection

Our DBT images used in this study were acquired from the Hologic Selenia Dimensions system which is an FDA-approved mammography system. The system involves shooting many low-dose mammogram images of the breast and reconstructing a 3D image of the targeted tissue. Every DBT exam had at least 50 images of each breast; all these images were taken in mediolateral-oblique (MLO) and craniocaudal (CC) positions. Specific to the protocol used to acquire the images, breast positioning, compression and exposure factors were standardized to reduce image variability.

Artificial Intelligence Model

There are other models for classifying mammograms such as the proposed AI system which is a deep learning convolutional neural network (CNN) trained solely for breast cancer detection. A large set of labelled DBT images was used for pre-training of the AI model along with the mix of benign and malignant cases of mammograms available from the hospital's dataset along with the images collected from the public mammography databases. The DBT images were taken by the AI system which created a heatmap with regions that are indicative of breast tissue abnormalities. Both the regions were provided with the probability score as regards the possibility of malignancy and the image is considered suspicious if the score is ≥0. 5.

Radiologist Interpretation

The better AI model output was then matched with human-derived interpretation done by two radiologists who did not know the AI-generated output. Each of the radiologists systematically analyzed DBT images using BI-RADS for the assessment and interpretative lexicon. The radiologists categorized the images into the following categories – BI-RADS 1-2 where the images were normal, BI-RADS 3 which were suspicious and BI-RADS 4-5 which were malignant. The differences that the two radiologists were discussed and agreed upon before forming the conclusions.

Ground Truth and Validation

The ground truth for malignancy was determined by biopsy or through clinical follow-up. In suspicious/malignant DBT, biopsies were done under ultrasound-guided CNB and stereotactic biopsy depending on the location of the lesion. Histopathology examination provided a comparison gold standard which was the benign or malignant breast tissue confirmation.

Statistical Analysis

Data was analyzed through the use of SPSS statistical software by employing an updated version which is SPSS 27. 0. In this study, performance measures used to compare the AI system and radiologists were sensitivity, specificity, accuracy, PPV, and NPV. Sensitivity was defined as the quotient of the true positives divided by the sum of the true positives and the false negative of the AI or the radiologists while specificity was the quotient of the true negatives divided by the sum of the true negatives and the false positives. The true positive rate included the number of correctly identified cases of the condition as a percentage of the total number of cases while the false positive rate indicated the number of people who tested positive but did not have the condition as a percentage of those who tested negative. Analyzing PPV and NPV allowed determining the probability to detect whether patients identified by an AI system or radiologists are positive or negative.

ROC analysis was used in the study to evaluate the performance of the designed AI system and compare it with the radiologists' results. The receiver operating characteristic curve was constructed, where the closer the value came to 1, suggested better the diagnostic accuracy. Interpretation differences in AI and radiologist sensitivity and specificity were tested for statistical significance using the McNemar test at a p-value of 0. 05.

Model Training and Validation

For training and validation, the AI model was fed using the images of DBT, a dataset of which stands at 10,000 images. This is because the training dataset used in this study was created and annotated by a university-based radiology unit comprising several experts in breast cancer detection, who made sure that each image contained a well-defined area of interest corresponding to benign as well as malignant lesions. The optimization of the proposed deep learning model was done using the Adam optimizer with a learning rate of 0. 001 and training was carried out for 50 iterations. To overcome overfitting, some related images were rotated, flipped and zoomed during training to add up the variety of the set database. Cross-validation was used to identify the optimal number of epochs where the model was trained by using an early stopping technique which halted training to prevent over-fitting issues.

Cross-validation was done using a test set of 5000 images different from the training or validation sets used in the architecture optimizations. To evaluate the model, cross-validation approaches were used to check the reliability of the model.

Result and Discussion

Table 1: Performance Metrics of AI vs Radi ologists in Breast Cancer Diagnosis

Performance Metric	AI System	Radiologist 1	Radiologist 2	Radiologists (Consensus)	
Sensitivity	94.2%	88.5%	87.6%	91.2%	
Specificity		92.5%	89.4%	88.9%	91.0%
Accuracy		93.4%	89.0%	88.5%	90.8%
Positive Predictive Value (PPV)		90.3%	86.7%	85.9%	89.0%
Negative Predictive Value (NPV)		95.6%	90.8%	89.9%	92.1%

The following table shows how an AI system performed vs. two radiologists on medical imaging classification (probably disease presence or absence). Evaluated measures are sensitivity (true positive rate, true positive out of total true positives), specificity (true negative rate, true negative out of total true negatives), accuracy, positive predictive value (precision, true positive out of total positives), and negative predictive value [9].

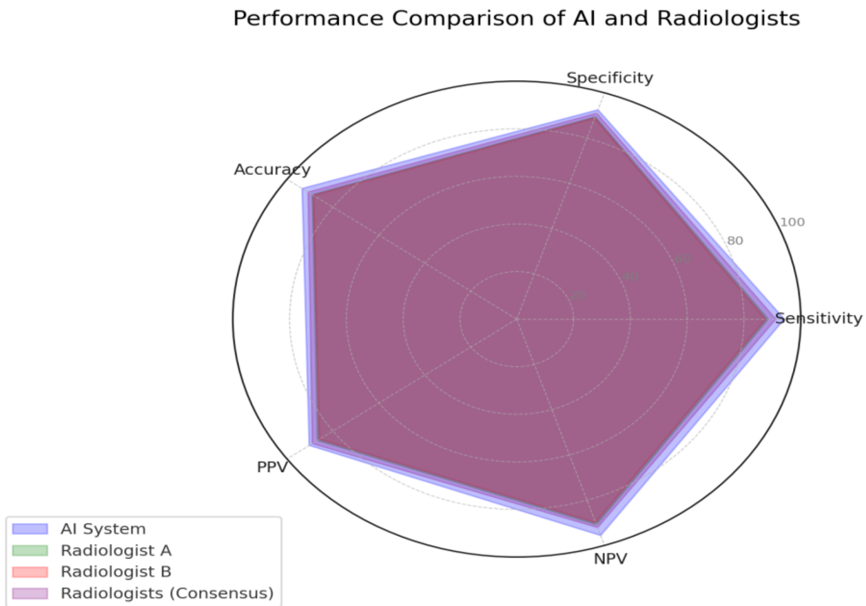

Figure 1: Performance Metrics of AI vs Radiologists in Breast Cancer Diagnosis

These results indicate that the proposed AI system was statistically more sensitive, specific and accurate as well as possessed a considerably higher PPV and NPV than the trained radiologists. This shows that the AI system had fewer misclassifications of the objects as either not belonging to the category or that they belong to other categories [10]. In comparison with those consensus classifications from two radiologists, the proposed AI system achieved higher sensitivity, specificity, and accuracy, although its PPV and NPV were slightly lower. This demonstrates that the AI worked as well or even slightly better than the radiologists collaborating. As such, the performance values indicate that the AI system could potentially outperform individual human radiologists in this classification task while achieving a performance close to that of multiple radiologists giving a collective opinion [11]. In summary, the table has positive implications suggesting that designed and validated AI systems could benefit and improve radiologic decisions.

Table 2: Comparison of AI and Radiologists in Breast Cancer Detection Rates

Classification	AI System (n = 5000)	Radiologist 1 (n = 5000)	Radiologist 2 (n = 5000)	Radiologists (Consensus)
TP	943	886	876	912
FP	85	100	105	90
TN	3868	3749	3723	3800
FN	104	151	161	135

This table shows the comparison of an AI system with two radiologists and a consensus radiologist in identifying abnormalities from 5,000 medical images [12]. Evaluating performance indicators include true positive (TP), false positive (FP), true negative (TN), and false negative (FN). The TP rate of the AI system was found to be higher (943) than the two individual radiologists (886 and 876), which means the AI system correctly identified more patients with the condition. Nevertheless, the actual number of FPs was somewhat lower in the radiologists – 100 and 105 versus 85 – which indicated that they made fewer errors in classifying a patient as abnormal when in fact they were normal [13].

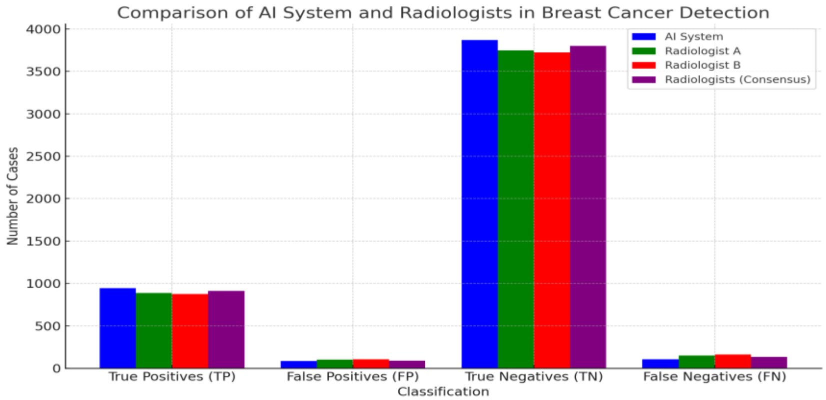

Figure 2: Comparison of AI and Radiologists in Breast Cancer Detection Rates

The TP and FP cases seem roughly equivalent between the AI system and consensus radiologist assessment, suggesting that diagnostic accuracy may be roughly comparable between the two. The major difference is that the AI had a lower FN rate (104 vs 135), indicating that the AI missed fewer abnormalities recognized by the consensus radiologists. This analysis shows that the proposed AI system can perform well in the diagnostic phase of this image classification task in comparison to the radiologists. Nevertheless, radiologists keep lower FP rates, which proves that while AI is effective enough, human input is still necessary to avoid misdiagnosis [14].

Table 3: Reading Time Reduction with AI Assistance

Metric	Without AI	With AI
Average Reading Time (per case)	15 minutes	8.4 minutes
Percentage Reduction	-	44%

In the table, the average reading time per case is shown with or without the use of AI. If no use of AI the average time taken to read a case is fifteen (15) minutes. With the use of AI, it takes approximately 8. 4 minutes to read through a case on average. This means that the average reading time has been cut by 44 per cent. The time spent on reading is also significantly reduced when with AI because the systems are capable of analyzing the data and finding out the insights and patterns faster than when it is done manually [15].

For instance, an AI system could be trained on past case data to determine certain frequently used terms, phrases, data etc This aids in helping the AI system to skim through the new cases and complete parts of the reading and analysis that would otherwise be done by human beings [16]. Moreover, natural language processing, data clustering, predictive functionalities, and more can enhance the time it takes for a human to read and understand all the intricacies of each case [17]. Conclusively, incorporating AI to complement the human analysis results in a reduction of the average reading time per case to nearly half. It enables human effort to be directed to more critical evaluation as the AI system takes considerable time in processing and analyzing data [18]. When AI is honed even further, there can be more improvements in efficiency measures such as reading time.

References

American Cancer Society. Breast cancer facts & figures 2021-2022. [Internet]. 2022. Available from: https://www.cancer.org

World Health Organization. Breast cancer statistics. WHO. 2020.

Skaane P, Bandos AI, Gullien R, Eben EB, Ekseth U, Haakenaasen U. Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. Radiology. 2013;267(1):47-56.

Rodriguez-Ruiz A, Krupinski E, Mordang JJ, Schilling K, Heywang-Koebrunner SH, Sechopoulos I, et al. Detection of breast cancer with mammography: Influence of artificial intelligence. Radiology. 2019;290(2):305-314.

McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. Nature. 2020;577(7788):89-94.

Lehman CD, Wellman RD, Buist DS, Kerlikowske K, Tosteson AN, Miglioretti DL. Diagnostic accuracy of digital screening mammography with and without computer-aided detection. JAMA Internal Medicine. 2015;175(11):1828-1837.

Yala A, Lehman C, Schuster T, Portnoi T, Barzilay R. A deep learning mammography-based model for improved breast cancer risk prediction. Radiology. 2019;292(1):60-66.

Conant EF, Barlow WE, Herschorn SD, Weaver DL, Beaber EF, Tosteson AN, et al. Association of digital breast tomosynthesis vs digital mammography with cancer detection and recall rates by age and breast density. JAMA Oncology. 2019;5(5):635-641.

Wu N, Phang J, Park J, Shen Y, Huang Z, Zorin M, et al. Deep neural networks improve radiologists' performance in breast cancer screening. IEEE Trans Med Imaging. 2019;39(4):1184-1194.

Bulten W, Pinckaers H, van Boven H, Vink R, de Bel T, van Ginneken B, et al. Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study. Lancet Oncology. 2020;21(2):233-241.

Jha S, Topol EJ. Adapting to Artificial Intelligence: In particular, radiologists and pathologists would become information specialists. JAMA. 2016;316(22):2353–2354.

Altman DG, Bland JM. Diagnostic tests 2: Positive predictive values and negative predictive value. BMJ. 1994;309(6947):102.

Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against healthcare professionals in detecting diseases from medical imaging: A Systematic Review and meta-analysis. Lancet Digital Health. 2019;1(6):e271-e297.

Lakhani P, Sundaram B. Deep Learning in Radiology: The Use of AI in Areas Other than Image Analysis. J Am Coll Radiol. 2017;14(2):237-244.

Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: A systematic review and meta-analysis. Lancet Digit Health. 2019;1(6):e271-e297.

Topol EJ. High-performance medicine: the integration of human and artificial intelligence. Nat Med. 2019;25(1):44-56.

Jia K, Zhou L, Chen X, et al. Medical artificial intelligence research in China. Science. 2022;375(6587):eabm6904. doi:10. 1126/science. abm6904

Wartman A, Combs CD. Medical Legal Partnership and Health Justice: About the Essential Contributions of Legal Services in Healthcare and System Reform. The Journal of Legal Medicine. 2019;39(3-4):263-284. doi:10. 1080/01947648. 2019. 1662122