Artificial Intelligence in Early Detection: Identifying Breast Cancer Before Clinical Diagnosis

Prasurjya Saikia*[1], Durgaprasad Kemisetti[2], Ananga Mohan Das[3], Charlisar Teron[4], Diptimonta Neog[5]

*1Assistant Professor, Faculty of Pharmaceutical Science, Himalayan University , Jollang, village, near central jail, Itanagar, Arunachal Pradesh 791111, India

2Associate Professor, Faculty of Pharmaceutical Science, Assam down town University, Sankar Madhab Path, Gandhi Nagar, Panikhaiti, Guwahati, Assam 781026, India

3Assistant Professor, Faculty of Pharmaceutical Science, Himalayan University , Jollang, village, near central jail, Itanagar, Arunachal Pradesh 791111, India

4Scholar, Faculty of Pharmaceutical Science, Assam down town University, Panikhaiti, Guwahati, Assam -781026, India.

5Department of Physics, North Eastern Regional Institute of Science and Technology, Nirjuli -791109, Arunachal Pradesh, India

CORRESPONDING AUTHOR-

Prasurjya Saikia

Email -prasurjyasaikia17@gmai.com

Phone no.995771237

ORCHID ID-

Prasurjya Saikia; 0009-0004-3785-5894

Durgaprasad Kemisetti; 0000-0003-2081-1794

Charlisar Teron; 0009-0002-8016-7718

Ananga Mohan Das -0009-0006-2328-1169

Diptimonta Neog-0000-0002-8690-0637

Abstract - Improving patient outcomes depends critically on early identification of breast cancer. In order to detect breast cancer up to five years before a clinical diagnosis, artificial intelligence (AI) has the potential to completely transform breast cancer screening. This paper examines this possibility. We explore the most recent developments in AI algorithms and how they relate to imaging in medicine, namely mammography. The paper looks at how AI can identify precancerous alterations that are invisible to the human eye by analysing minute patterns in breast tissue. We go over the difficulties and possibilities in creating and evaluating AI models for early detection, including model interpretability, data quality, and ethical issues. The ultimate goal of this analysis is to demonstrate how artificial intelligence (AI) has the potential to drastically lower breast cancer mortality by enabling much earlier detection.

Keywords-*Artificial Intelligence, Breast Cancer, Personalized medicine,Digital Mammography*

1. Introduction to AI in Healthcare : For many years, intelligent computer systems have played a big role in society. Research and development on artificial intelligence (AI) is of interest to many different areas of the economy, government, industry, technology, healthcare, and security and defence.[1] The use of AI is currently changing in many disciplines due to the convergence of innovative AI approaches, enormous computer processing power, and the ubiquitous increase of digital data gathering and storage, especially in science and health. AI systems are being created, investigated, and tested for illness detection, prognostication, and as support tools for clinical decision-making in the field of cancer care, much as in other healthcare domains. In particular, current research on breast cancer involves an international endeavour to create sophisticated machine learning algorithms for analysing screening mammograms, which might enhance breast cancer screening by decreasing false positives.[2]

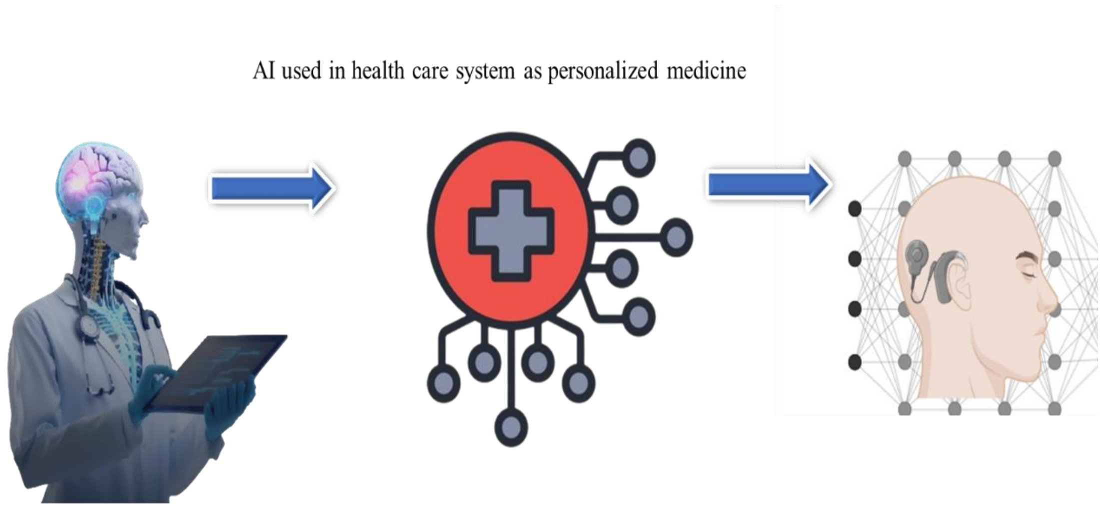

Fig :1 – Respresentation of AI in health care system

AI has applications in breast cancer diagnostics that extend beyond imaging to include pathology result interpretation. In sentinel lymph node biopsies, for example, AI has aided in the identification of metastatic breast cancer in whole-slide pictures. In order to comprehend how contemporary AI systems might enhance screening procedures, we centre our efforts in this work on the early identification of breast cancer.[3] As the most common disease to be diagnosed, breast cancer (BC) has overtaken lung cancer, with around three million cases and 700,000 fatalities recorded in

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

2020.[4] The annual number of instances of breast cancer has been increasing since the mid-2000s. Nonetheless, there has been a significant decline in the fatality rate from breast cancer.

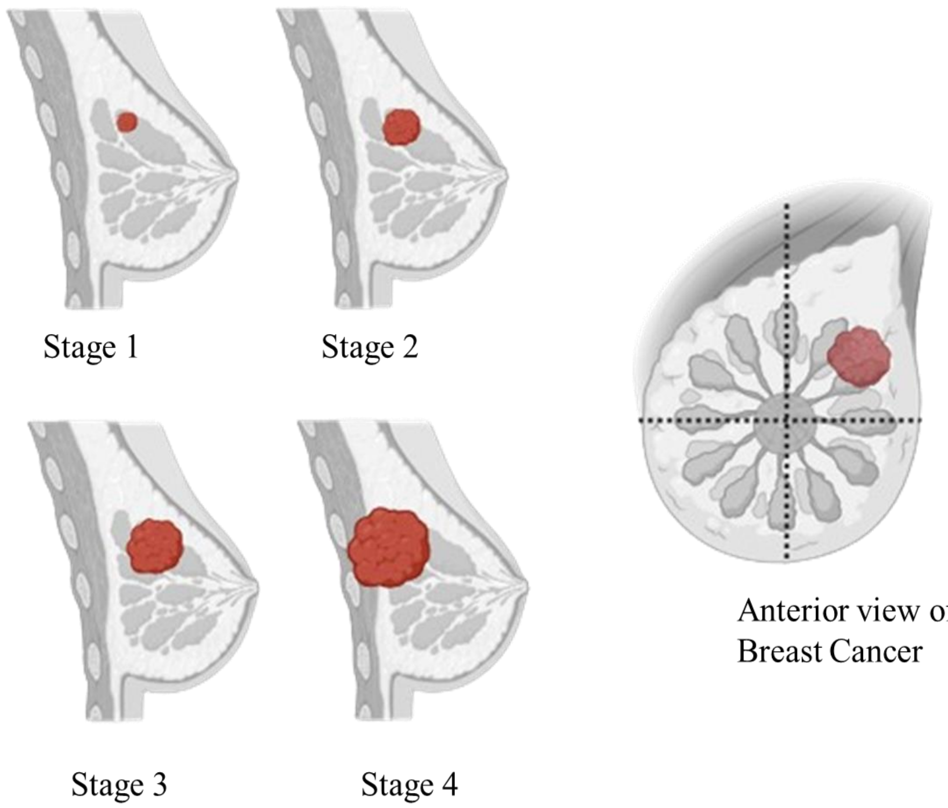

Fig :2 – Representation of stages of Breast Cancer

The tendency differs according to ethnicity. For example, while Non-Hispanic White (NHW) people have greater rates of breast cancer than Non-Hispanic Black (NHB) people, NHW people's mortality rate has fallen more than NHB's. Because of this, NHB people have higher mortality rates even if their incidence rates are greater among NHW.[5] Many screening programs are based on age, which is the most often researched risk factor for breast cancer, but the illness can also develop as a result of a number of other variables. High body mass index, a history of hyperplastic or neoplastic breast illness, and a family history of breast cancer are important risk factors. Long-term oestrogen exposure, whether from early menstruation or late menopause, is another important risk factor.[6] Many screening programs are based on age, which is the most often researched risk factor for breast cancer, but the illness can also develop as a result of a number of other variables. High body mass index, a history of hyperplastic or neoplastic breast illness, and a family history of breast cancer are important risk factors. Long-term oestrogen exposure, whether from early menstruation or late menopause, is another important risk factor.[7]

2. Literature review -We conducted a literature search from 2010 to 2018 to review advances in AI methods for breast cancer detection. The review focused on studies assessing AI approaches in screening, including those with quantitative performance data compared to established standards. Eligible studies evaluated AI in breast cancer screening among women, not restricted to those with cancer or biopsy history. Excluded were studies on simulated lesions, those without performance data, or those with fewer than 100 subjects or 200 images. Commentary, editorial articles, reviews, and conference abstracts were also excluded.

3. **Artificial intelligence is transforming healthcare, particularly in diagnostic processes-**Physicians using artificial intelligence (AI) to interpret pictures and diagnose patients more accurately can save lives.[8] Machine learning (ML) and deep learning (DL) are the two primary subfields of artificial intelligence (AI). With machine learning (ML), computers may discover patterns and insights by learning from data and making predictions without explicit programming. Beyond machine learning (ML), deep learning (DL) learns directly from unprocessed data, doing away with the requirement for human-defined characteristics. This makes it especially effective for deciphering complicated data patterns and enables more automatic and sophisticated learning processes. AI offers enormous promise for imaging in medicine. It can help radiologists analyse results, optimise imaging systems, and automatically identify lesions in different organs. AI-driven CAD systems improve diagnosis and prognosis by offering second viewpoints. AI also improves picture post-processing operations such as segmentation and registration.[9]

4. **Current Methods of Breast Cancer Detection-**

International approaches for breast cancer screening differ. Women who want to get screened either self-refer or through their doctors are recommended to institutions in the US and other similar nations. Institutions differ in their ways of screening and interpretation of images. On the other hand, a lot of European nations have government-run screening programs that use uniform tools and processes. Women enrolled in these programs are periodically invited for screenings, and those who uncover anything abnormal are referred to hospitals for additional assessment.

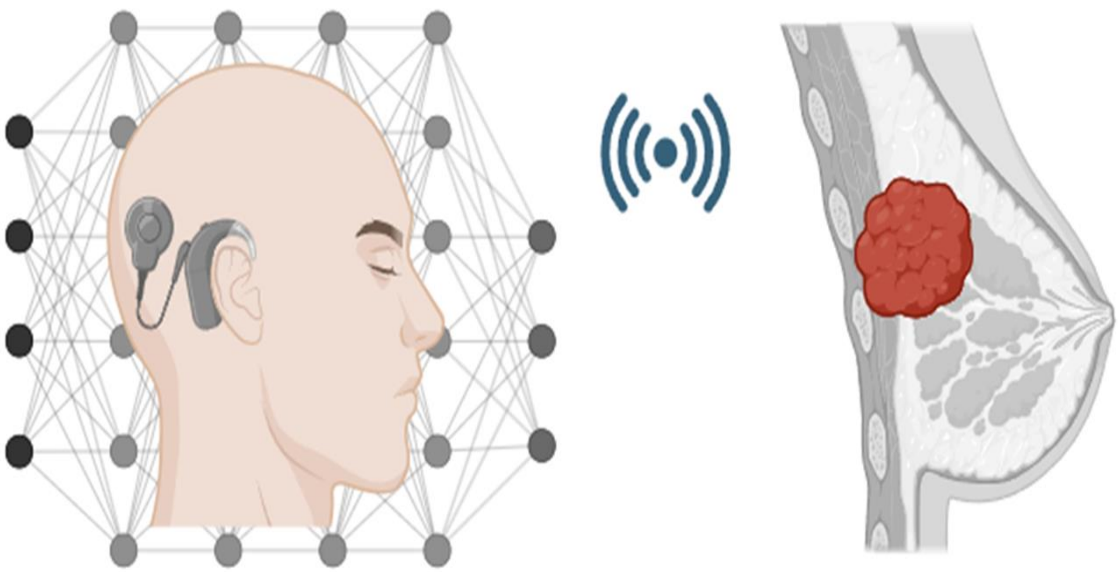

Fig : 3– Representation of Current methods used in the detection of Breast cancer

a) **Digital Mammography-**Breast cancer screening practices vary globally. In countries like the US, screening is often institution-based, with women self-referring or being referred by their doctors. Screening methods and image interpretation vary between institutions. In contrast, many European countries have government-run screening programs with standardized procedures and equipment. Women in these programs receive periodic invitations for screening, and suspicious findings lead to referrals to hospitals for further evaluation.[10] Compared to traditional film mammography, digital mammography (DM) has a number of advantages, one of which is a more effective workflow. For the general population, DM is largely similar to film mammography; however, in certain patient populations, DM has demonstrated better performance. Furthermore, the development of sophisticated image processing tools and advanced imaging procedures like tomosynthesis and specific breast CT is made possible by digital technology.[11] Whether film-based or digital, mammography is limited by its two-dimensional representation of a three-dimensional breast. This presents a significant challenge of tissue superposition, where overlapping structures obscure underlying abnormalities. To address this, standard mammography protocols involve taking two orthogonal views of each breast: the mediolateral oblique (MLO) and cranio-caudal (CC) projections. Radiologists carefully compare these images to identify potential lesions, taking into account factors like lesion shape, density, margins, and calcifications. Nevertheless, the limitations of 2D imaging frequently lead to missed cancers, especially in dense breast tissue where malignant lesions can be obscured by overlapping glandular tissue.

b) **Digital breast tomosynthesis-**Over the last 20 years, the development and clinical acceptance of digital breast tomosynthesis (DBT) have been driven by the intrinsic constraints of standard 2D mammography. By taking numerous pictures of the breast from various angles, DBT successfully reduces the overlapping tissue artefact that is a problem with traditional mammography and provides a pseudo-3D perspective of the breast. As a result, as compared to digital mammography (DM), DBT has shown increased cancer detection rates and typically lower recall rates. This is especially true in situations where the initial DM recall rate was high.[12] Because there are more image slices in DBT images than in standard mammograms, interpreting DBT images takes a lot longer. This is a significant obstacle to the widespread use of DBT in screening programs. Automated techniques are essential to address this. By accelerating navigation through the image stack and aiding in lesion detection, AI can dramatically reduce radiologists' workload. Furthermore, AI can help standardize interpretation, potentially improving the consistency of DBT's impact on cancer detection rates across different studies.

Conventional techniques used in detecting breast cancer -Computer-Aided Detection (CADe) and Computer-Aided Diagnosis (CADx) are the two main categories that resulted from the transition from film to digital mammography. CADe is concerned with identifying possible abnormalities, like masses or clusters of microcalcifications. Traditional CADe systems work in three steps: first, the image is preprocessed to enhance suspicious features; second, the system identifies potential areas of interest; and third, these regions are evaluated to determine the likelihood of a true abnormality. Finally, these regions are evaluated to determine the likelihood of a true abnormality, and those exceeding a certain threshold are flagged for radiologist review.[13] In comparison to CADe, CADx systems evaluate the possibility that a lesion found is benign or malignant. Though CADx and the last phase of CADe are comparable, CADx does not use a simple threshold.Conventional CADe and CADx systems rely on pre-programmed features, which teach computers what to search for in the form of particular traits connected to lesions that seem worrisome. This strategy is very different from contemporary AI-based techniques. Researchers have created algorithms that examine both breast images in order to replicate the human interpretation process and improve CADe performance. Comparing similar characteristics between the same and various breasts is one way to do this.[14]

5. Limitations and challenges of these conventional - Because various algorithms currently employ different testing data and methodologies, it is difficult to assess their efficacy. Nor are there many independent studies that examine their efficacy.[15] High-quality, labelled, representative data, such as those on the distribution of pathologies, demographics, and breast density, are lacking. Because most of the research datasets available now are limited and originate from a single institution or mammography machine vendor, overfitting of algorithms may occur. In order to facilitate independent testing with fresh datasets and prevent bias from recurring situations, the data used for training, validating, and testing AI systems should be public. Since ground truth has a major influence on the algorithm, it must be dependable. Potential biases must be taken into account before implementing AI in clinical practice. An example of such a bias is the "anchoring effect." This occurs when an image has indicators indicating potential cancer, leading viewers to overrely on these indicators and skewing their judgement.[16]

6. AI Algorithms and Technology- Studies compared the performance of deep learning CNNs to the state-of-the-art, such as conventional CADe/CADx, in the early stages of breast cancer diagnosis in mammography. For instance, in order to identify masses in DBT, Fotin et al. compared deep learning CNNs with standard CADe. They discovered that using deep learning improved the sensitivity for suspicious lesions from 83.2% to 89.3% and for malignant lesions from 85.2% to 93.0%. [17]Instead of using traditional CADe, Becker et al. conducted another early research where they tested the effectiveness of deep learning versus radiologists. They employed a commercial deep learning-based image analysis method that is not authorised for use in medicine and was designed for industrial application. Two distinct datasets were used for training and testing the algorithm: one had an equal mixture of 50% malignant and 50% control instances, while the other, which more closely mirrored real screening settings, had around 10% malignant and 90% control cases. The study shown that deep learning algorithms, especially ones intended for non-medical imagery, can be trained to identify breast cancer in DM, even though the second group still had a substantially higher cancer prevalence than actual screenings. Two of the three readers outperformed the algorithm by a large margin in the high prevalence group, while the algorithm was about as good as the radiologists in the lower prevalence set.[18]

7. Benefits of Early Detection- There are several advantages to early illness identification, particularly with cancer, which enhance patient outcomes. Early diagnosis typically translates into easier treatment and management, which increases the likelihood of survival. Less harsh therapies are possible with early identification, which lessens the patients' physical and psychological stress. Additionally, by avoiding costly treatments required for severe illnesses, it reduces healthcare expenses. Furthermore, prompt treatment can halt the disease's progression, enhancing the lives of affected individuals and their family. Better long-term health and well-being are supported by early diagnosis, which identifies health problems before they worsen.[19]

8. Future Prospects-

Future developments in imaging technology, such as more sophisticated CT and MRI scanners and the creation of liquid biopsies, will transform early illness identification and treatment by offering less intrusive screening alternatives and better pictures. AI will be essential in the medical field because it can swiftly and correctly analyse enormous volumes of data, spot trends and abnormalities, and customise treatment regimens based on a patient's genetic composition and medical background. When paired with wearable medical technology, AI-powered solutions will increase diagnostic precision and allow for real-time patient monitoring and early illness diagnosis, which will result in prompt treatments and better patient outcomes.

9. Current state-of-art AI algorithms for digital mammography and digital breast tomosynthesis- CNNs are now the foundation of the greatest AI algorithms for identifying and categorising lesions in mammography and DBT. There are a number of businesses have FDA-approved, CE-marked, or in the last stages of review commercial AI applications. In order to assess the efficacy of a commercial AI system for DM, Rodriguez Ruiz and colleagues gathered nine data sets from US and European locations. The probability-of-malignancy (PoM) ratings and DM pictures from studies comparing DM with other imaging modalities were included in these data sets. The radiologists performed breast screening. The final data set included 2,652 examinations from four separate suppliers, including 653 instances of cancer. 28,296 separate interpretations of the photos were produced from the interpretations of 101 radiologists. At the case level, the AI system's performance was statistically no worse than the average performance of the 101 radiologists[20] The study discovered that the AI outperformed 61% of the radiologists when comparing the AI system's performance to that of individual radiologists. The ROC curves demonstrate that, regardless of whether previous tests were taken into account for evaluation, the AI's performance was consistently comparable to that of the radiologists.This extensive study's strength is found in the diversity of the data as well as the vast number of instances and readers. Images from systems manufactured by four distinct vendors were included in the research, along with radiologists' interpretations from seven different nations, including the US and Europe, which have varied screening practices. This study, however, did not entirely resolve all issues and does not adequately forecast how AI would do in actual breast cancer screenings in comparison to humans. others of the data sets were unilateral, others eliminated previous pictures, and all were enhanced and evaluated in a laboratory environment. Radiologists frequently rely on the ability to compare recent photos with older ones, but AI systems are not yet capable of doing this. Notwithstanding these drawbacks, the radiologists' performance was not appreciably impacted in comparison to the AI, even when previous pictures were accessible. Determining any bias in the study's results may be challenging due to factors like lab settings and non-screening illness prevalence that may have affected the findings. Thus, while this thorough analysis offered insightful information regarding the level of AI assessment in DM at the moment, more research is required, ideally with large-scale datasets to evaluate radiologist performance in real screens. Very large training datasets are needed for whole-image classification techniques that are not trained with annotated pictures. Furthermore, these approaches require additional procedures to indicate to the user where worrisome discoveries are located. For example, Kim et al. trained, validated, and tested a deep learning CNN that could classify images as malignant or not and produce heat maps indicating the areas that contributed most to the final classification decision using a dataset of over 4,000 cancer cases and nearly 25,000 normal cases, all without pixel-level annotations.[21]

10. Graphical Abstract –

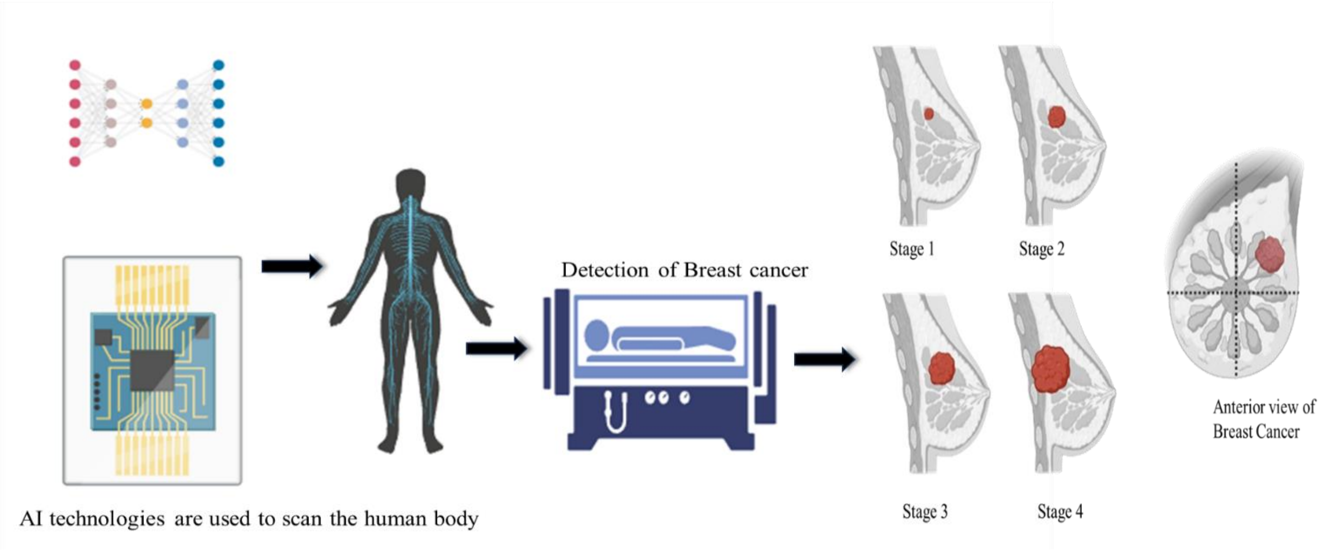

Fig:4- Representation of Graphical Abstract of AI detection of Breast Cancer

11. Conclusion-

Computers are becoming more adept at identifying breast cancer in mammograms, or images of the breast. Even while they seem very promising, additional research is still needed to determine exactly how effectively they function.Scientists used images from various locations and medical professionals to test AI. Even while the AI performed rather well, we are unsure if it can detect cancer as accurately as human doctors can. This is due to the fact that the computer was taught through unique images that were not found in typical examinations. Furthermore, physicians routinely use AI to detect changes in patient records by comparing recent images to older ones.To truly understand how AI functions, we need to test it on a large number of standard images. If we succeed in doing this and maintain making the AI smarter, it could help doctors find cancer earlier and save lives.

12. References

1) Plan, Strategic. "The national artificial intelligence research and development strategic plan." *National Science and Technology Council, Networking and Information Technology Research and Development Subcommittee* (2016).
2) Houssami, Nehmat, Christoph I. Lee, Diana SM Buist, and Dacheng Tao. "Artificial intelligence for breast cancer screening: opportunity or hype?." *The Breast* 36 (2017): 31-33.
3) Wang, Dayong, Aditya Khosla, Rishab Gargeya, Humayun Irshad, and Andrew H. Beck. "Deep learning for identifying metastatic breast cancer." *arXiv preprint arXiv:1606.05718* (2016).
4) Sung, Hyuna, Jacques Ferlay, Rebecca L. Siegel, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, and Freddie Bray. "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries." *CA: a cancer journal for clinicians* 71, no. 3 (2021): 209-249.
5) Martini, Rachel, Lisa Newman, and Melissa Davis. "Breast cancer disparities in outcomes; unmasking biological determinants associated with racial and genetic diversity." *Clinical & experimental metastasis* 39, no. 1 (2022): 7-14.
6) Pike, Malcolm C., M. D. Krailo, B. E. Henderson, J. T. Casagrande, and D. G. Hoel. "'Hormonal'risk factors,'breast tissue age'and the age-incidence of breast cancer." *Nature* 303, no. 5920 (1983): 767-770.
7) Jakes, R. W., S. W. Duffy, F. C. Ng, F. Gao, and E. H. Ng. "Mammographic parenchymal patterns and risk of breast cancer at and after a prevalence screen in Singaporean women." *International journal of epidemiology* 29, no. 1 (2000): 11-19.
8) Stoitsis, John, Ioannis Valavanis, Stavroula G. Mougiakakou, Spyretta Golemati, Alexandra Nikita, and Konstantina S. Nikita. "Computer aided diagnosis based on medical image processing and artificial intelligence methods." *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 569, no. 2 (2006): 591-595.
9) Choy, Garry, Omid Khalilzadeh, Mark Michalski, Synho Do, Anthony E. Samir, Oleg S. Pianykh, J. Raymond Geis, Pari V. Pandharipande, James A. Brink, and Keith J. Dreyer. "Current applications and future impact of machine learning in radiology." *Radiology* 288, no. 2 (2018): 318-328.
10) Thomassin-Naggara, Isabelle, Corinne Balleyguier, Luc Ceugnart, Patrice Heid, Greg Lenczner, Aurélien Maire, Brigitte Séradour, Laurent Verzaux, and Patrice Taourel. "Artificial intelligence and breast screening: French Radiology Community position paper." *Diagnostic and Interventional Imaging* 100, no. 10 (2019): 553-566.
11) Skaane, P. "Studies comparing screen-film mammography and full-field digital mammography in breast cancer screening: updated review." *Acta radiologica* 50, no. 1 (2009): 3-14.
12) Ciatto, Stefano, Nehmat Houssami, Daniela Bernardi, Francesca Caumo, Marco Pellegrini, Silvia Brunelli, Paola Tuttobene et al. "Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study." *The lancet oncology* 14, no. 7 (2013): 583-589.

13) Ganesan, Karthikeyan, U. Rajendra Acharya, Chua Kuang Chua, Lim Choo Min, K. Thomas Abraham, and Kwan-Hoong Ng. "Computer-aided breast cancer detection using mammograms: a review." *IEEE Reviews in biomedical engineering* 6 (2012): 77-98.

14) Van Engeland, Saskia, and Nico Karssemeijer. "Combining two mammographic projections in a computer aided mass detection method." *Medical Physics* 34, no. 3 (2007): 898-905.

15) Topol, Eric J. "High-performance medicine: the convergence of human and artificial intelligence." *Nature medicine* 25, no. 1 (2019): 44-56.

16) Becker, Anton S., Magda Marcon, Soleen Ghafoor, Moritz C. Wurnig, Thomas Frauenfelder, and Andreas Boss. "Deep learning in mammography: diagnostic accuracy of a multipurpose image analysis software in the detection of breast cancer." *Investigative radiology* 52, no. 7 (2017): 434-440.

17) Bruno, Michael A., Eric A. Walker, and Hani H. Abujudeh. "Understanding and confronting our mistakes: the epidemiology of error in radiology and strategies for error reduction." *Radiographics* 35, no. 6 (2015): 1668-1676.

18) Kooi, Thijs, Geert Litjens, Bram Van Ginneken, Albert Gubern-Mérida, Clara I. Sánchez, Ritse Mann, Ard den Heeten, and Nico Karssemeijer. "Large scale deep learning for computer aided detection of mammographic lesions." *Medical image analysis* 35 (2017): 303-312.

19) Shaikh, Khalid, Sabitha Krishnan, and Rohit M. Thanki. *Artificial intelligence in breast cancer early detection and diagnosis*. Cham: Springer, 2021.

20) Rodriguez-Ruiz, Alejandro, Kristina Lång, Albert Gubern-Merida, Mireille Broeders, Gisella Gennaro, Paola Clauser, Thomas H. Helbich et al. "Stand-alone artificial intelligence for breast cancer detection in mammography: comparison with 101 radiologists." *JNCI: Journal of the National Cancer Institute* 111, no. 9 (2019): 916-922.

21) Kim, Eun-Kyung, Hyo-Eun Kim, Kyunghwa Han, Bong Joo Kang, Yu-Mee Sohn, Ok Hee Woo, and Chan Wha Lee. "Applying data-driven imaging biomarker in mammography for breast cancer screening: preliminary study." *Scientific reports* 8, no. 1 (2018): 2762.