Deep Learning Model to Predict the Risk of Developing Breast Cancer in Mammography Based – A Pilot Study in Southern Brazil

Manuela Zereu[1] Carmela Farias da Silva[1] Katsuki Arima Tiscoski[1] Roberta de Almeida da Silva[1] Thiago Krieger[1] Márcio Gustavo Gusmão Scherer[2] Liana Rigon Rojas Lima[2] Jeison Philomena Molina[2] Letícia Batistela[2] Débora Roesler[2] Juliano Maurício Bordignon[2] Maria Claudia Schardosim Cotta de Souza[3] Antônio Nocchi Kalil[1]

[1] Research and Teaching, Hospital Santa Casa de Porto Alegre, Porto Alegre, RS, Brazil
[2] Technology and Innovation, Empresa Pública de Tecnologia da Informação e Comunicação da Prefeitura de Porto Alegre (PROCEMPA), Porto Alegre, RS, Brazil
[3] Department of Public Health, Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA), Porto Alegre, RS, Brazil

Address for correspondence Manuela Zereu, MD, PhD, Santa Casa de Porto Alegre Hospital, Rua Professor Annes Dias, 295, CEP 90020-090, Porto Alegre, RS, Brazil (e-mail: manuelazereu@gmail.com).

Braz J Oncol 2026;22:s00451814169.

Abstract

Introduction Breast cancer is the leading cause of cancer-related deaths among women in Brazil, highlighting the importance of early detection to improve outcomes. Artificial intelligence (AI) has garnered considerable attention for its potential to enhance breast cancer screening by reducing unnecessary exams, minimizing diagnostic errors, and increasing efficiency and accuracy—exemplified by advanced tools like Mirai.

Materials and Methods The present retrospective study analyzed 1,000 patients who underwent bilateral mammography from December 2019 to April 2024 at Hospital Santa Casa de Porto Alegre. All mammograms extracted in digital imaging and communications in medicine (DICOM) format were anonymized and processed by the Mirai algorithm to generate risk scores. Predictive performance was evaluated using discrimination metrics, such as C-index and area under the curve (AUC), as well as threshold analyses (F1-score, Youden's J) to estimate cancer risk.

Results Mirai obtained a C-index of 0.76 (95% CI: 0.72–0.80) and an AUC of 0.81. The analysis further evaluated the F1-score and Youden's J statistic in an attempt to establish risk thresholds for cancer development.

Conclusion The results suggest that the Mirai model holds promise as a valuable tool for breast cancer detection, particularly for early identification of high-risk patients.

Keywords
- artificial intelligence
- deep learning
- breast neoplasms
- early diagnosis

Introduction

Cancer is among the leading causes of death worldwide, with breast cancer being the most prevalent type among women globally and in Brazil. This disease poses a significant public health challenge due to its high incidence, mortality, and morbidity rates.[1,2]

According to the most recent evidence, annual mammographic screening is recommended for women aged 40 to 74 years, as it is associated with earlier diagnosis, better

received
May 9, 2025
accepted
October 4, 2025

DOI https://doi.org/10.1055/s-0045-1814169.
ISSN 2526-8732.

Thieme Revinter Publicações Ltda., Rua Rego Freitas, 175, loja 1, República, São Paulo, SP, CEP 01220-010, Brazil

prognosis, less aggressive treatments, and reduced mortality and morbidity. For women aged 75 years and older, screening should be considered only in those with a life expectancy greater than 7 years. In younger women or in different screening intervals, the balance between benefits and risks becomes less favorable, and screening should be individualized based on risk factors.[3]

Systematic reviews of randomized trials over the past 50 years have shown that mammographic screening for women aged 50 to 70 reduces breast cancer mortality risk. However, these findings mainly reflect older mammography techniques and do not account for current treatment protocols, which may limit their relevance. A 2016 systematic review found moderate-quality evidence that mammography reduces breast cancer mortality risk in women aged 50 to 59 years (RR: 0.86, 95% CI: 0.68–0.97, 7 trials), with a more significant reduction in women aged 60 to 69 years (RR: 0.67, 95% CI: 0.54–0.83, 5 trials).[4]

To improve screening strategies, artificial intelligence (AI) has gained increasing prominence in healthcare, particularly for its ability to detect very small lesions beyond the sensitivity of conventional diagnostic tools, support clinical decision-making through predictive scores, reduce unnecessary invasive procedures and related adverse outcomes.[5,6]

In this context, Yala et al.[7] developed a deep learning-based risk model, named Mirai (Jameel Clinic), to predict future breast cancer risk solely from mammographic images, overcoming limitations of traditional clinical models. Trained on over 210,000 mammograms and externally validated in some institutions, Mirai demonstrated high accuracy (AUC 0.76–0.79, at 5 years) and outperformed conventional models such as Tyrer-Cuzick (MagView). The model identifies high-risk patients not detected by traditional methods and shows potential for individualized risk stratification, improving screening programs and enabling personalized preventive interventions.[7]

Despite that, external validations need to be done in several populations with diverse clinical and environmental characteristics. Therefore, we validated this AI model using mammograms, with and without a cancer diagnosis, to predict the risk of breast cancer over a period of 5 years in a cohort of women from southern Brazil.

Materials and Methods

The present retrospective observational study selected from medical records 1,000 women who had a bilateral screening mammography examination (2-dimensional digital mammography) between December 2019 and April 2024 at Hospital Santa Casa de Porto Alegre. The data obtained from medical records were birth date, date of the first mammography, and date of the last mammography.

The inclusion criteria were female patients, between 20 and 90-years-old, who underwent at least 2 mammography exams with positive and negative diagnoses for breast cancer.

The exclusion criteria were patients with only one mammography exam.

The authors provided justification for the waiver of the informed consent form.

Procedures

Hospital Santa Casa de Porto Alegre

The researchers accessed the medical records of eligible patients and selected mammography exams through the CareStream system. Once a patient's exam was selected, the following steps were performed:

a) Downloading the digital imaging and communications in medicine (DICOM) images format standard of mammography: left craniocaudal (L-CC), left mediolateral-oblique (L-MLO), right craniocaudal (R-CC), and right mediolateral-oblique (R-MLO).
b) Images were anonymized through software developed by Empresa Pública de Tecnologia da Informação e Comunicação da Prefeitura de Porto Alegre (PROCEMPA, "Porto Alegre Company of Public IT and Communications"), which removes all information that could identify the patient.
c) Using the same software, the anonymized image was sent to a PROCEMPA S3 storage, with access via login and password.
d) An Excel (Microsoft Corp.) table was filled with the coding of each patient and the information mentioned above. A flag was placed on patients who developed cancer.
e) All data were sent to PROCEMPA using Transport Layer Security (TLS).

Empresa Pública de Tecnologia da Informação e Comunicação da Prefeitura de Porto Alegre (PROCEMPA)

This company has developed an anonymization software to facilitate the secure transfer of medical imaging exams in DICOM format from the infrastructure of the Hospital Santa Casa to its own. Considering that these exams contain sensitive patient information, the anonymization process was strategically designed to ensure data privacy and compliance with regulations governing the handling of personal health information. By implementing this solution, the sensitive metadata associated with the exams is automatically removed or replaced with nonidentifiable attributes before leaving the hospital's infrastructure, requiring no additional effort from the personnel selecting the images.

After the anonymized mammography images and the Excel table containing patient data are transferred to PROCEMPA's infrastructure, they are securely stored in an Amazon S3 (Amazon.com, Inc.) compatible storage solution within PROCEMPA's private datacenter. Following storage, the subsequent steps are taken:

a) The OncologIA software developed by PROCEMPA retrieves the anonymized mammography images from the S3 storage.
b) For each patient, OncologIA collects all four standard DICOM images (L-CC, L-MLO, R-CC, R-MLO) and constructs a Web service (WS) payload, which is then used

as input data for Mirai's WS (the API that performs the inference against the AI model), which is running as a Docker container within PROCEMPA's infrastructure.[8]

c) After the WS's response, Mirai's output (prediction scores) and associated data (Excel table [Microsoft Corp.]) are stored in a dedicated database. This centralized repository facilitates the organization and accessibility of results for subsequent statistical analysis and performance evaluation of the Mirai model.

d) The stored results are used to evaluate the performance of the Mirai model.

The C-index and the area under the receiver operating characteristic (ROC) curve (AUC) for 1 to 5 years outcomes were calculated to evaluate the overall performance of the Mirai model, providing an aggregate measure of its ability to distinguish positive and negative cases.

Additionally, key metrics such as the F1-score and Youden's J statistic were calculated to determine the optimal thresholds for prediction cancer risk. These statistical evaluations provide critical insights into the model's effectiveness and guide further refinement to improve its predictive capabilities. The F1-score threshold prioritizes a balance between precision and recall, aiming to minimize false positives and negatives. On the other hand, Youden's J aims to maximize the model's overall effectiveness by identifying the threshold that optimally balances sensitivity and specificity. This approach enhances the model's ability to distinguish between patients with and without breast cancer while minimizing misclassification.

The current study was approved by the Santa Casa de Porto Alegre's Research Ethics Committee by number 6.676.546. The authors provided justification for the waiver of the informed consent form.

Results

The total sample comprised 1,000 female patients, of whom underwent two bilateral mammograms. For this study, the earliest available mammogram for each patient was selected. From the selected mammogram, all standard views were obtained, including L-CC, L-MLO, R-CC, and R-MLO. It took 1,000 exams to train the AI Mirai in our population.

A total of 1,000 patients aged 20 to 90 years were included to achieve the sample size required for appropriate calibration of the AI Mirai in our population. Notably, over 85% of participants were between 40 and 75-years-old, as illustrated in ►**Fig. 1**.

Among these patients, 174 (17.4%) were already known to have breast cancer at the time of the study, with mammograms that presented breast imaging reporting and data system (BI-RADS, American College of Radiology) category 6, confirmed by prior biopsy. This group was essential for assessing the model's ability to accurately predict breast cancer in patients with confirmed diagnoses. Additionally, 743 (74.3%) patients did not develop breast cancer and formed the control group used to evaluate the model's specificity and false-positive rates (BI-RADS 2; routine screening). The remaining 83 (8.3%) individuals were in remission from breast cancer. The inclusion of this last subgroup contributed not only by assessing the model's capacity to differentiate between patients in remission and those at risk of developing the disease but also by evaluating its ability to identify patients with an active cancer diagnosis.

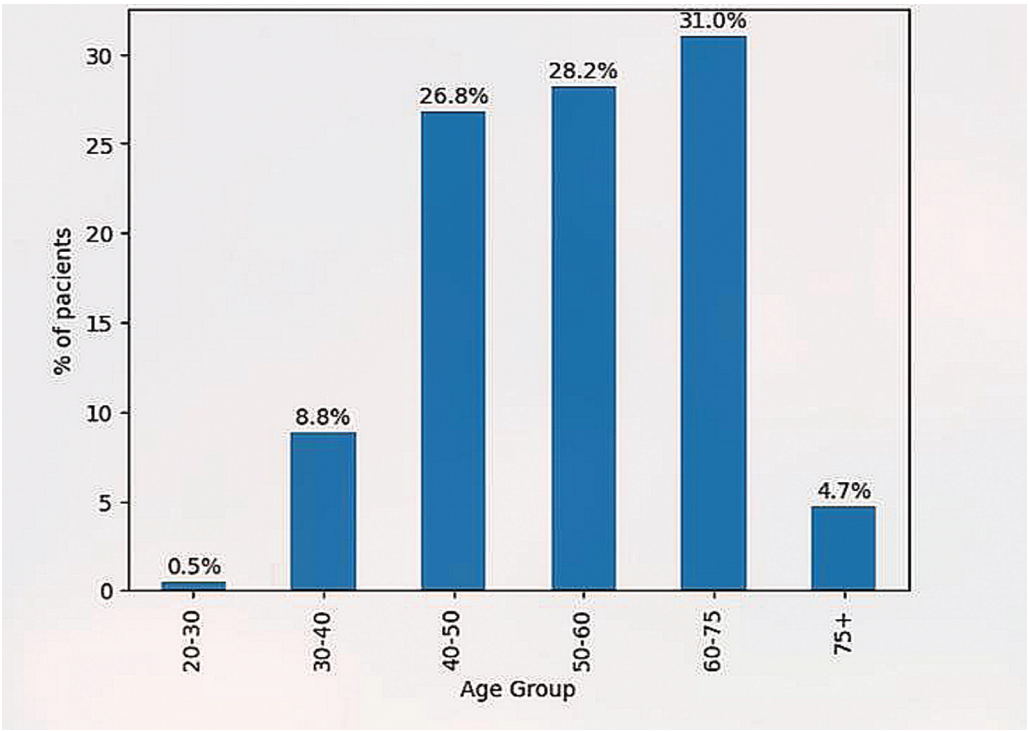

Fig. 1 Patients distribution by age group.

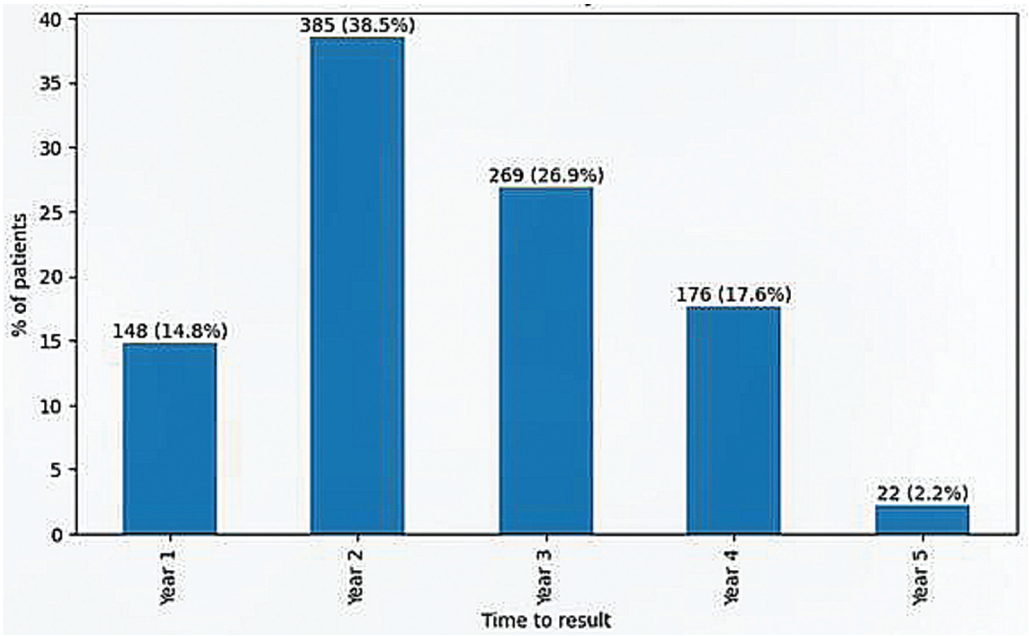

Fig. 2 Distribution of patients who underwent mammography over 5 years (total sample).

This diverse group provided a robust foundation for validating the model's predictive performance across a range of clinical scenarios.

►**Fig. 2** illustrates the distribution of patients who underwent mammography over the 5-year follow-up period. In the 1st year (2019–2020), 148 patients (14.8%) underwent mammography. The 2nd year recorded the highest number of exams, with 385 patients (38.5%). In the 3rd year, 269 patients (26.9%) underwent mammography, followed by 176 patients (17.6%) in the 4th year. In the 5th and final year, only 22 patients (2.2%) underwent the exam.

Also, the risk of developing breast cancer within 5 years was analyzed across the entire sample (►**Fig. 3**), with the distribution of patients according to the interval between their initial exam and follow-up at which they were diagnosed. The data shows that 14.8% of patients were diagnosed within 1 year, 38.5% after 2 years, 26.9% after 3 years, 17.6% after 4 years, and only 2.2% after 5 years.

The median number of mammograms performed per individual included in the study is 3. The range of mammograms performed per individual ranges from 2 (minimum) to 9 (maximum).

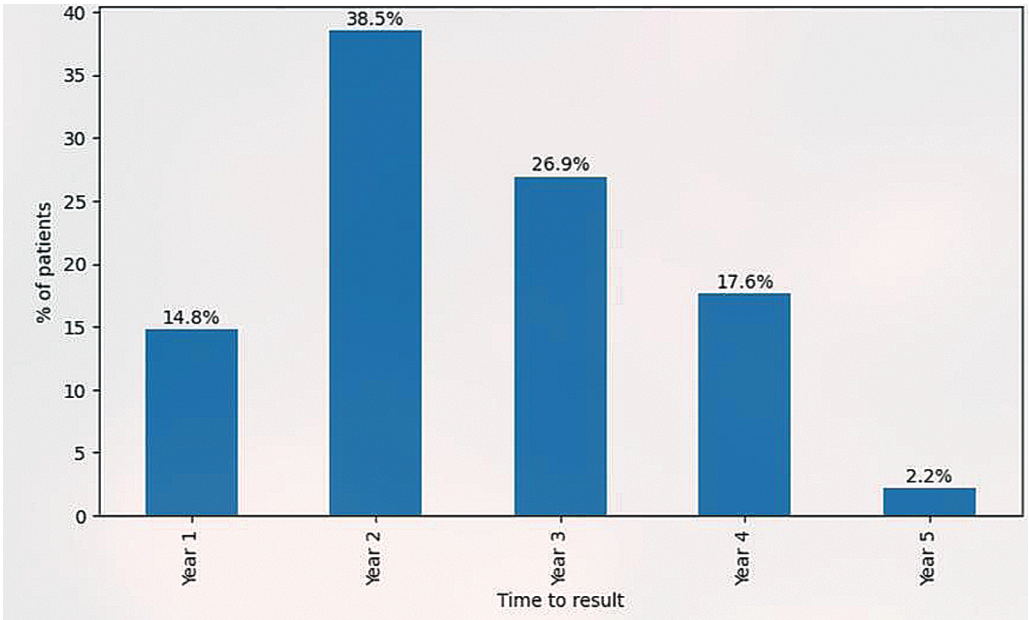

Fig. 3 Risk of developing breast cancer within 5 years.

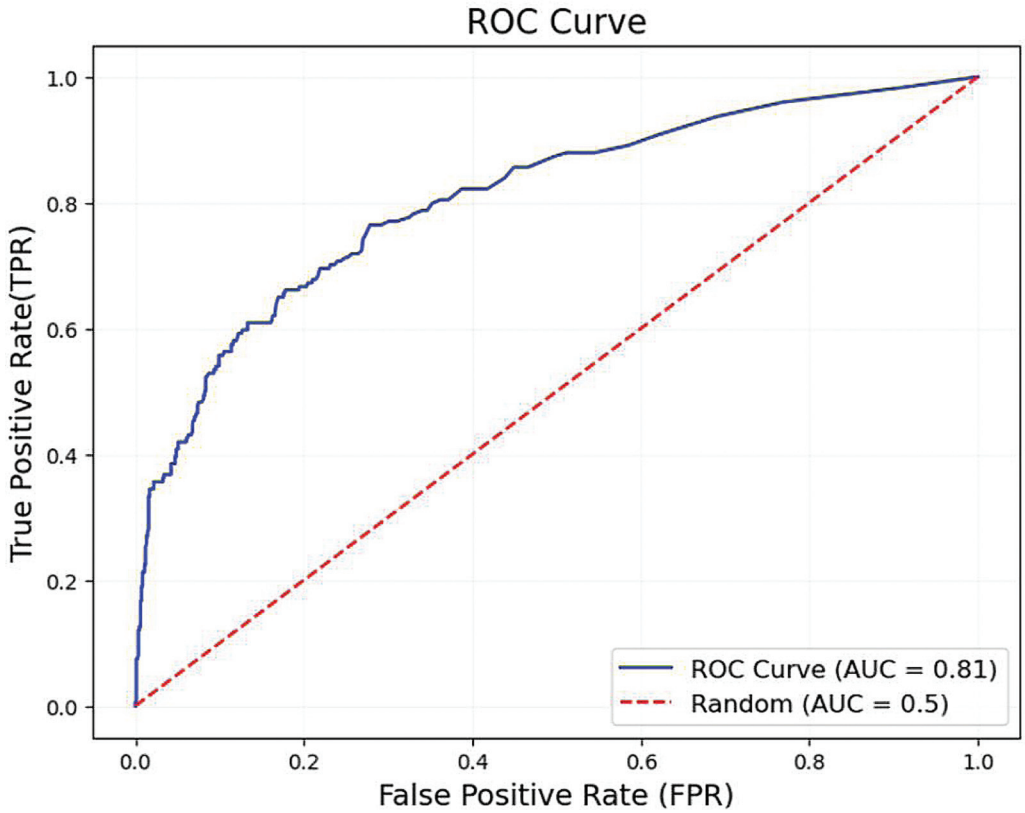

Fig. 4 Performance of the Mirai deep learning model for breast cancer prediction.

According to the AUC evaluation, Mirai achieved a 0.81 value (►**Fig. 4**). The C-index obtained was 0.76 (95% CI: 0.72–0.80).

The selection of an appropriate threshold to evaluate risk is critical in determining the performance and clinical applicability of predictive models like Mirai. In this study, analyzing the 1st year, we evaluate 2 potential thresholds derived from the F1-score (0.0159) and Youden's J statistic (0.0038).

Discussion

Although the greatest benefit of breast cancer screening has been reported in women aged 40 to 75 years, our inclusion criteria encompassed patients aged 20 to 90 years in order to reach the target sample size required for proper calibration of the AI Mirai in our population. Importantly, over 85% of the sample was concentrated within the 40-to-75 age range, which aligns with the population most likely to benefit from

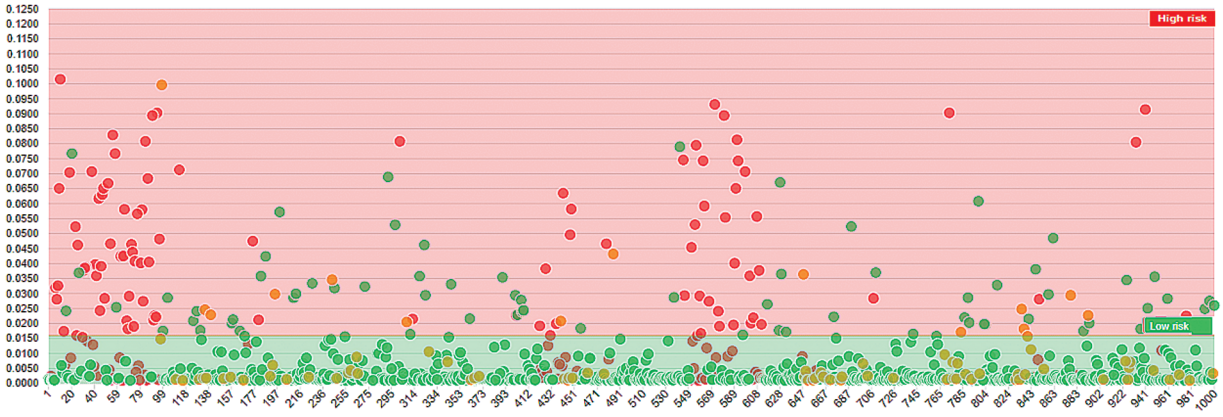

Fig. 5 Patient risk classification based on F1-score threshold (red dots: patients with cancer; green dots: patients without cancer; orange dots: patients undergoing treatment).

screening. This broader age inclusion, while necessary from a methodological standpoint, may introduce heterogeneity. However, the predominance of participants in the clinically relevant age group supports the robustness and applicability of our findings.

According to a study by Swets,[9] an AUC between 0.80 and 0.90 is considered excellent in many medical applications. The optimal value can vary depending on the specific clinical context and the consequences of false positives and false negatives.[9] Our result (0.81) was similar to the study of Yala et al.,[7] indicating strong discriminative power. Therefore, this AI deep learning model demonstrated promising predictive capabilities for breast cancer.

It is crucial to acknowledge that relying solely on these statistical measures for risk threshold determination may not be sufficient. As highlighted by Lamb et al., clinical context and patient-specific factors play a significant role in risk assessment. They compared the Mirai risk assessment model to traditional risk models, demonstrating the potential of deep learning models to improve risk stratification for supplemental screening with magnetic resonance imaging (MRI).[10]

The choice between these thresholds has significant clinical implications. In this research, the F1-score threshold (0.0159) offers a lower false positive rate (9.93%) and identifies 55.74% of patients who developed breast cancer. However, the Youden's J threshold (0.0038), while increasing sensitivity (76.43% of cancers detected), comes at the cost of a substantially higher false positive rate (27.85%). ►**Fig. 5** graphically presents the AI inference results (Y-axis) for each patient (X-axis). The threshold that separates the layers (red for high risk and green for low risk) was determined based on the F1-score. In the graph, red points represent patients who developed cancer, green those who did not, and orange those undergoing treatment.

The F1-score threshold, with a lower false positive rate, may be preferable in contexts where minimizing unnecessary follow-ups is a priority. Conversely, the Youden's J threshold could be more suitable in situations where higher sensitivity is needed, such as when early detection is paramount and resources for follow-up are readily available. Ultimately, the choice of threshold should be informed by the clinical context, resource availability, and the institution's tolerance for the trade-offs between sensitivity and specificity. Another critical consideration when selecting a threshold is the institution's capacity and willingness to manage false positives, as these generate additional diagnostic procedures, treatments, and tests, all of which incur costs and may burden the healthcare system.

The clinical significance of Mirai is further highlighted by its ability to identify patients at significantly higher risk of breast cancer. During the first year of analysis, the AI model's inference results (where a higher value indicates a higher risk) were, on average, more than five times higher for patients who developed breast cancer compared to those who did not (0.028 vs. 0.005, respectively). This significant difference underscores Mirai's potential to effectively stratify patients based on their individual risk levels, enabling more targeted and personalized preventative or therapeutic interventions. Despite that, dense breast tissue limits the sensitivity of mammography exams, which could impact the Mirai performance.[11]

As described in the *Methods* section, we did not collect clinical data from the patients recruited for the study. Several authors support the ability of the AI Mirai to estimate breast cancer risk solely from mammographic images, without requiring additional clinical information. Yala et al.[7] pioneered a model that predicts future breast cancer risk based exclusively on mammographic images, overcoming the limitations of traditional models that rely on demographic and laboratory data. Similarly, Lehman et al.[12] showed that AI tools applied directly to mammograms can enable personalized and efficient screening strategies without the need for supplemental clinical data. clinical data. Lamb et al.,[10] further demonstrated Mirai's accuracy by comparing predictions based solely on digital mammograms with those of traditional models. Omoleye et al.[13] reinforced this capability showing that Mirai performs robustly even in populations with limited access to specialized care, where clinical data collection can be challenging. Arasu et al.[14] and Park et al.[15] also confirmed Mirai's superior performance over models based on clinical information, validating its effectiveness exclusively with digital mammograms and in diverse population contexts. Collectively, these studies indicate that Mirai AI represents a significant advance, enabling individualized and accurate risk stratification through mammographic image analysis alone.

Regarding cost-effectiveness, Hill et al.[16] showed a monetary benefit to the UK National Health Service (NHS) when using Mirai in the breast cancer screening program, stratified by risk, based on quality-adjusted life-years (QALYs), compared to the current screening program. Specific risk-stratified breast cancer screening regimens (RSBCRs), determined by risk thresholds, have shown annual economic benefits to the NHS that can reach £60.4 million (US$77.3 million) per year.[16]

Avendano et al.[17] also evaluated Mirai in a cohort of 3,110 Mexican women. Their analysis showed that Mirai achieved a C-index of 0.63 (95% CI: 0.6–0.7). In their study, a Mirai index score $>$ 0.029 (10% threshold) was used to identify high-risk individuals. The authors reported that individuals in the high-risk group had nearly three times the risk of developing breast cancer compared to those in the low-risk one.[17]

This pilot study has some limitations. First, because it focused on validating Mirai AI within a specific population, relevant medical data for correlating with the findings were not collected. However, previous studies have demonstrated that, unlike traditional clinical risk models such as Tyrer-Cuzick and the Gail Model (National Institute of Health), this

algorithm estimates breast cancer risk solely from mammographic images, without requiring additional clinical or demographic information, as described above. This is the main advantage of AI Mirai.

Second, the study's retrospective design and small sample size of mammography images represent a limitation, although the AUC and C-Index results were consistent with two larger retrospective studies.[10,12] Third, 5-year follow-up data were not available for all patients included in the study. Fourth, while we used AUC to assess the AI model's predictive performance, relying solely on this metric may not be sufficient for comprehensive validation. However, our findings demonstrated strong discriminatory performance, comparable to other studies.

Conclusion

The Mirai AI tool demonstrated promising capabilities in predicting which patients are at high risk of developing breast cancer within the next 5 years. The C-index value was similar to that reported in the study by Yala et al.,[7] reinforcing the model's consistency. Additionally, the high AUC demonstrated here indicates strong predictive performance. This AI algorithm has the potential to complement existing breast cancer risk guidelines and enhance experts' interpretation. The results of this study, including the calculated F1-score and Youden's J thresholds, will be instrumental in designing a future clinical trial to prospectively validate the Mirai model and establish clinically relevant risk thresholds.

Authors' Contributions

MZ: project administration, supervision, writing – review & editing; CFS: writing – original draft, writing – review & editing, conceptualization, resources; KAT: writing – original draft, writing – review & editing, formal analysis; RAS: conceptualization, writing – review & editing; TK: conceptualization, writing – review & editing; MGGS: conceptualization, formal analysis, methodology, software, writing – original draft, writing – review & editing; LRRL: conceptualization, writing – review & editing; JPM: formal analysis, methodology, software, writing – review & editing; LB: conceptualization, software; writing – review & editing; DR: conceptualization, software; writing – review & editing; JMB: formal analysis, methodology, software, writing – review & editing; MCSCS: conceptualization, formal analysis, writing – review & editing; ANK: conceptualization, investigation, project administration, writing – review & editing.

Funding

The authors declare that they did not receive funding from agencies in the public, private or nonprofit sectors to conduct the present study.

Conflict of Interests

The authors have no conflict of interests to declare.

References

1 Ministério da Saúde; Instituto Nacional de Câncer (INCA). Controle do Câncer de Mama –Dados e Números – Incidência: Apresenta dados de incidência do câncer de mama no Brasil, regiões e estados. [cited December 26, 2024]. Ministério da Saúde. Available from: https://www.gov.br/inca/pt-br/assuntos/gestor-e-profissional-de-saude/controle-do-cancer-de-mama/dados-e-numeros/incidencia

2 World Health Organization. Newsroom – Fact sheets – Detail – Breast cancer. [cited December 26, 2024].Available from: https://www.who.int/news-room/fact-sheets/detail/breast-cancer

3 Urban LABD, Chala LF, Paula IB, et al. Recommendations for breast cancer screening in Brazil, from the Brazilian College of Radiology and Diagnostic Imaging, the Brazilian Society of Mastology, and the Brazilian Federation of Gynecology and Obstetrics Associations. Radiol Bras 2023;56(04):207–214. Doi: 10.1590/0100-3984.2023.0064-en

4 Nelson HD, Pappas M, Cantor A, Griffin J, Daeges M, Humphrey L. Harms of Breast Cancer Screening: Systematic Review to Update the 2009 U.S. Preventive Services Task Force Recommendation. Ann Intern Med 2016;164(04):256–267. Doi: 10.7326/M15-0970

5 Rana M, Bhushan M. Machine learning and deep learning approach for medical image analysis: diagnosis to detection. Multimedia Tools Appl 2022;82:26731–26769. Doi: 10.1007/s11042-022-14305-w

6 Santeramo R, Damiani C, Wei J, Montana G, Brentnall AR. Are better AI algorithms for breast cancer detection also better at predicting risk? A paired case-control study. Breast Cancer Res 2024;26(01):25. Doi: 10.1186/s13058-024-01775-z

7 Yala A, Mikhael PG, Strand F, et al. Toward robust mammography-based models for breast cancer risk. Sci Transl Med 2021;13(578): eaba4373. Doi: 10.1126/scitranslmed.aba4373

8 Docker. Docker: The Industry-Leading Container Runtime. [cited January 23, 2025]. Docker Inc. Available from: https://www.docker.com/products/container-runtime/

9 Swets JA. Measuring the accuracy of diagnostic systems. Science 1988;240(4857):1285–1293. Doi: 10.1126/science.3287615

10 Lamb LR, Mercaldo SF, Ghaderi K, Carney A, Lehman CD. Comparison of the Diagnostic Accuracy of Mammogram-based Deep Learning and Traditional Breast Cancer Risk Models in Patients Who Underwent Supplemental Screening with MRI. Radiology 2023;308(03):e223077. Doi: 10.1148/radiol.223077

11 Jin Z, Zhang S, Zhang L, Chen Q, Liu S, Zhang B. Artificial Intelligence Risk Model (Mirai) Delivers Robust Generalization and Outperforms Tyrer-Cuzick Guidelines in Breast Cancer Screening. J Clin Oncol 2022;40(20):2280–2281. Doi: 10.1200/JCO.21.02908

12 Lehman CD, Mercaldo S, Lamb LR, et al. Deep Learning vs Traditional Breast Cancer Risk Models to Support Risk-Based Mammography Screening. J Natl Cancer Inst 2022;114(10): 1355–1363. Doi: 10.1093/jnci/djac142

13 Omoleye OJ, Woodard AE, Howard FM, et al. External Evaluation of a Mammography-based Deep Learning Model for Predicting Breast Cancer in an Ethnically Diverse Population. Radiol Artif Intell 2023;5(06):e220299. Doi: 10.1148/ryai.220299

14 Arasu VA, Habel LA, Achacoso NS, et al. Comparison of Mammography AI Algorithms with a Clinical Risk Model for 5-year Breast Cancer Risk Prediction: An Observational Study. Radiology 2023; 307(05):e222733. Doi: 10.1148/radiol.222733 PubMed

15 Park EK, Lee H, Kim M, et al. Artificial Intelligence-Powered Imaging Biomarker Based on Mammography for Breast Cancer Risk Prediction. Diagnostics (Basel) 2024;14(12):1212. Doi: 10.3390/diagnostics14121212

16 Hill H, Roadevin C, Duffy S, Mandrik O, Brentnall A. Cost-Effectiveness of AI for Risk-Stratified Breast Cancer Screening. JAMA Netw Open 2024;7(09):e2431715. Doi: 10.1001/jamanetworkopen.2024.31715

17 Avendano D, Marino MA, Bosques-Palomo BA, et al. Validation of the Mirai model for predicting breast cancer risk in Mexican women. Insights Imaging 2024;15(01):244. Doi: 10.1186/s13244-024-01808-3

