Original research

Predicting cardiovascular events from routine mammograms using machine learning

Jennifer Yvonne Barraclough ,[1,2] Ziba Gandomkar,[3] Robert A Fletcher,[1] Sebastiano Barbieri ,[4,5] Nicholas I-Hsien Kuo ,[4] Anthony Rodgers,[6] Kirsty Douglas,[7] Katrina K Poppe ,[8] Mark Woodward ,[9,10] Blanca Gallego Luxan,[4] Bruce Neal,[9,11] Louisa Jorm ,[4] Patrick Brennan,[3] Clare Arnott[1,2]

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/heartjnl-2025-325705).

For numbered affiliations see end of article.

Correspondence to
Dr Jennifer Yvonne Barraclough; Jbarraclough@georgeinstitute.org.au

JYB and ZG are joint first authors.

Received 8 January 2025
Accepted 15 June 2025

ABSTRACT

Background Cardiovascular risk is underassessed in women. Many women undergo screening mammography in midlife when the risk of cardiovascular disease rises. Mammographic features such as breast arterial calcification and tissue density are associated with cardiovascular risk. We developed and tested a deep learning algorithm for cardiovascular risk prediction based on routine mammography images.

Methods Lifepool is a cohort of women with at least one screening mammogram linked to hospitalisation and death databases. A deep learning model based on DeepSurv architecture was developed to predict major cardiovascular events from mammography images. Model performance was compared against standard risk prediction models using the concordance index, comparative to the Harrells C-statistic.

Results There were 49 196 women included, with a median follow-up of 8.8 years (IQR 7.7–10.6), among whom 3392 experienced a first major cardiovascular event. The DeepSurv model using mammography features and participant age had a concordance index of 0.72 (95% CI 0.71 to 0.73), with similar performance to modern models containing age and clinical variables including the New Zealand 'PREDICT' tool and the American Heart Association 'PREVENT' equations.

Conclusions A deep learning algorithm based on only mammographic features and age predicted cardiovascular risk with performance comparable to traditional cardiovascular risk equations. Risk assessments based on mammography may be a novel opportunity for improving cardiovascular risk screening in women.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Cardiovascular screening tools are poorly used in women, with the WHO highlighting the need for novel approaches to cardiovascular risk prediction.

⇒ Existing algorithms based on breast arterial calcification (BAC) from mammographic images have been developed for cardiovascular risk prediction; however, there are limitations to the utility of BAC alone as a cardiovascular risk prediction tool.

WHAT THIS STUDY ADDS

⇒ This is the first deep learning model incorporating all breast characteristics/architecture from routine screening mammograms, as opposed to BAC alone, to be developed for major adverse cardiovascular event risk prediction.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ The potential impact of our research lies in using existing infrastructure and mammography screening pathways, already used by women, to predict cardiovascular risk, using a 'two for one' automated process.

Linked

► http://dx.doi.org/10.1136/heartjnl-2025-326657

Check for updates

To cite: Barraclough JY, Gandomkar Z, Fletcher RA, *et al*. *Heart* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/heartjnl-2025-325705

INTRODUCTION

Cardiovascular disease results in 18 million deaths annually and contributes to significant morbidity and reduced health-related quality of life.[1] Cardiovascular disease and its risk factors are underrecognised and undertreated in women, and risk prediction algorithms have underperformed in women,[2] adversely impacting outcomes.[3] For example, Australian women are 12% less likely than men to have cardiovascular risk factors assessed in primary care,[3] and in Australia in 2020, only 49% of eligible women had sufficient risk factors recorded to enable a cardiovascular risk assessment.[4] Despite newer developed risk scores, PREVENT[5] and Q4[6] performing better in women than men, they require extensive medical data to be accurate. No Australian programmes currently target gender inequity in cardiovascular risk screening, highlighting the need for novel approaches to identify 'at-risk' women.

In Australia, and globally, breast screening mammography is offered to women free-of-charge through a national screening programme, Breast Screen Australia, and 50% of eligible women aged 50–74 years attend screening biannually.[7] In addition to early cancer diagnosis, mammograms offer information about cardiovascular risk. Breast arterial calcification (BAC) has been shown to correlate with the risk of cardiovascular events[8] and with vascular risk factors such as diabetes, hypertension and hypercholesterolaemia.[9] However, BAC is not associated with obesity and is inversely associated

with smoking, suggesting using BAC alone to predict cardiovascular risk may have limitations.[8] Other mammographic features, including microcalcifications and breast density, have also been shown to be associated with cardiometabolic disease risk and mortality but are yet to be evaluated together.[10 11]

In addition to persisting uncertainties in the observational epidemiology, other challenges to the use of mammography for cardiovascular risk prediction have been the absence of standardised assessments, time-consuming nature of manual evaluations and inter-reader variability. Even the automated quantification of BAC has failed to overcome the differences in breast composition.[12] Fully automated deep learning analysis of whole breast architecture/characteristics, rather than BAC alone, may be more accurate at predicting cardiovascular events, but has not previously been investigated.

This study aims to derive and internally validate an algorithm that predicts cardiovascular risk in women attending routine screening mammography for breast cancer, assessing the potential for using mammography as a 'two for one' screening test.

METHODS

Recruitment

This study included women enrolled 2009–2020 in the Lifepool cohort registry, living in Victoria, Australia. The Lifepool cohort was established by the Peter MacCallum Cancer Centre, the University of Melbourne, and the Royal Melbourne Hospital, with participants recruited from Breast Screen sites across metropolitan and rural areas. It was established in 2009 and has enrolled 54 000 women. At enrolment, participants completed a baseline health survey and consented to the registry accessing the images from any screening mammography undertaken at Breast Screen and to link to their routinely collected health data through the Victorian Admitted Episodes Database for hospital admissions and the National Death Index. Further information on the Lifepool Cohort Study is available at www.lifepool.org.

Participants

Eligibility of Lifepool participants for this study was based on there being at least one set of screening mammogram images available and no hospitalisation for a cardiovascular cause recorded in the Victorian Admitted Episodes Database before their first mammogram.

Mammography measures

The first available set of right and left breast digital mammographic images was used for each participant. Wherever possible, image data for mammograms recorded on two different occasions were used. However, in 9% of cases, only one set of mammographic images was recorded prior to the first cardiovascular event, and in those cases, it was used in isolation.

Other exposure measures

The baseline health survey completed at enrolment captured self-reported data including age, smoking status, alcohol intake, body mass index, diabetes history and use of antihypertensive, lipid lowering and antiplatelet therapies. Additional data included menopause, parous history and use of hormonal therapy, and factors potentially altering breast architecture, such as radiation therapy, breast surgery and breast cancer.

Outcomes

The primary outcome was 'Extended Major Adverse Cardiovascular Events' (extended MACE) defined as either (1) a hospitalisation event recorded in the Victorian Admitted Episodes Database with a primary or secondary diagnosis encompassing an International Classification of Disease version 10 Australian Modification code for atherosclerotic cardiovascular disease or heart failure (online supplemental appendix table 1) or (2) a corresponding death record in the National Death Index. Secondary outcomes were the individual components of Extended MACE: myocardial infarction, stroke, atherosclerotic disease, heart failure and cardiovascular death.

ANALYSIS

Baseline characteristics were summarised as means and proportions and tabulated. Participant follow-up was from the date of their first mammogram to the date of their first recorded outcome or the censoring date defined by the last available linkage to the National Death Index on 31 December 2020. Analyses were conducted using R V.4.2.3.

Cardiovascular risk prediction models: We used an image encoder to extract mammography features, which were combined in a neural-network-based model to predict risk of extended MACE over a 10-year time horizon (figure 1). The encoder included two submodules for mapping the input images into a radiomic signature: one for extracting deep-learnt features of the mammography images and one for extracting 196 conventional radiomic features (online supplemental appendix page 5 app 2–3 sub-module 2). Prior to deep feature extraction, images were normalised to reduce the internal covariate shift, improve training stability and harmonise across imaging sites and equipment. However, normalisation may obscure subtle features like faint or small calcifications. To address this, conventional

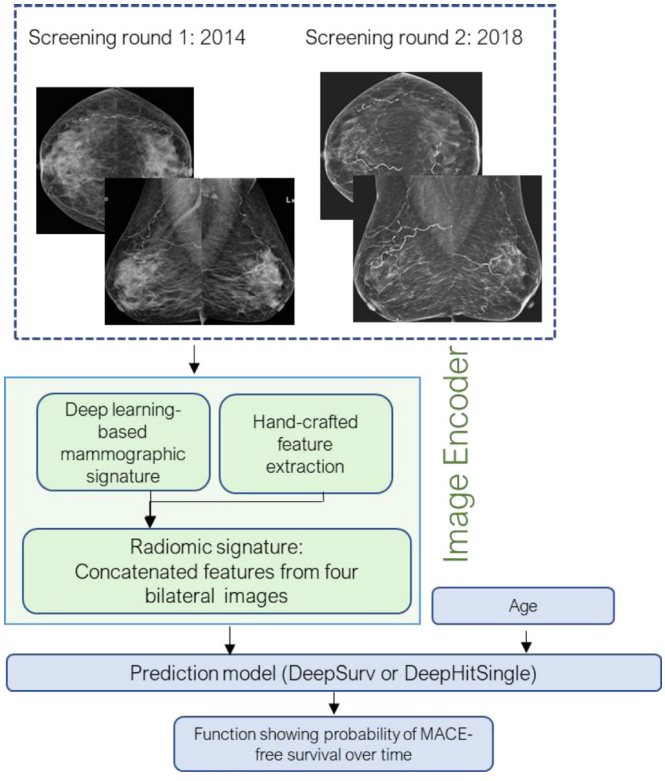


Figure 1 Model for predicting cardiovascular risk using mammography and age. Radiomic model for risk prediction fed into the image encoder using one or two screening mammography rounds and predicting MACE-free survival using DeepSurv and DeepHitSingle prediction models. MACE, major adverse cardiovascular event.

radiomic features were extracted from original, non-normalised images to preserve such details and complement deep features. These included first-order statistics, texture features (GLCM, GLRLM, GLSZM), higher-order matrices (NGTDM, SFM) and transform-based features (Gabor, RFS, Fourier), following validated protocols in breast cancer imaging. Because some scanners contributed few cases, deep learning may have under-represented their feature space. Including handcrafted radiomics mitigated this limitation by enhancing feature diversity across devices. A sensitivity analysis (online supplemental table 6) showed a small but statistically significant performance gain with their inclusion.

The final radiomic signature was obtained by concatenating the deep learning-based and conventional features from the images (figure 1). Where two image sets were available, the time interval (in years) between the two sets of images used for each individual was included as a feature in the radiomic signature to explore whether this was associated with model performance. The radiomic signature was then combined with the participant's age at the baseline survey to predict the risk of extended MACE risk using DeepSurv, a semiparametric non-linear continuous-time model.[13] The model was trained and tested using nested cross-validation. Tuning hyperparameters of DeepSurv and DeepHitSingle models was undertaken to improve the performance of the deep learning-based model.[14 15] A list of the hyperparameter search spaces for each deep learning model is provided in the online supplemental table 2. Figure 2 shows the nested cross-validation method, including the details of training of the model.[16] This method tested three different sets of inputs:

1. Mammography model—based on age and radiomic data.
2. Clinical model—based on clinical characteristics (All variables in table 1) without radiomic data.
3. Combined model—based on clinical characteristics and radiomic data.

The analyses were repeated using DeepHitSingle, a non-parametric non-linear discrete-time model as a sensitivity analysis[17] (online supplemental appendix and online supplemental table 2).

All models were built in Python.

Comparison of performance between models: We evaluated the performance of the deep learning-based survival models using the time-dependent concordance index, the integrated inverse probability of censoring weighted Brier score, and the integrated inverse probability of censoring weighted binomial log-likelihood.[15] The performance metrics are further described in the online supplemental appendix.

The calibration of the models was compared by dividing the participants into deciles based on their predicted risk of experiencing an event in the next 10 years and the mean predicted risk for each group from each model was plotted against the observed risk for the group calculated using Kaplan-Meier survival methods. The slopes of the regression lines between the predicted and observed risks for each model were then compared (online supplemental figure 1).

Comparison of performance against standard risk prediction models: Current cardiovascular risk prediction models use a Cox proportional hazards model as their underlying framework and usually report Harrell's C-index as a measure of discriminative power. We tabulated key features of established cardiovascular

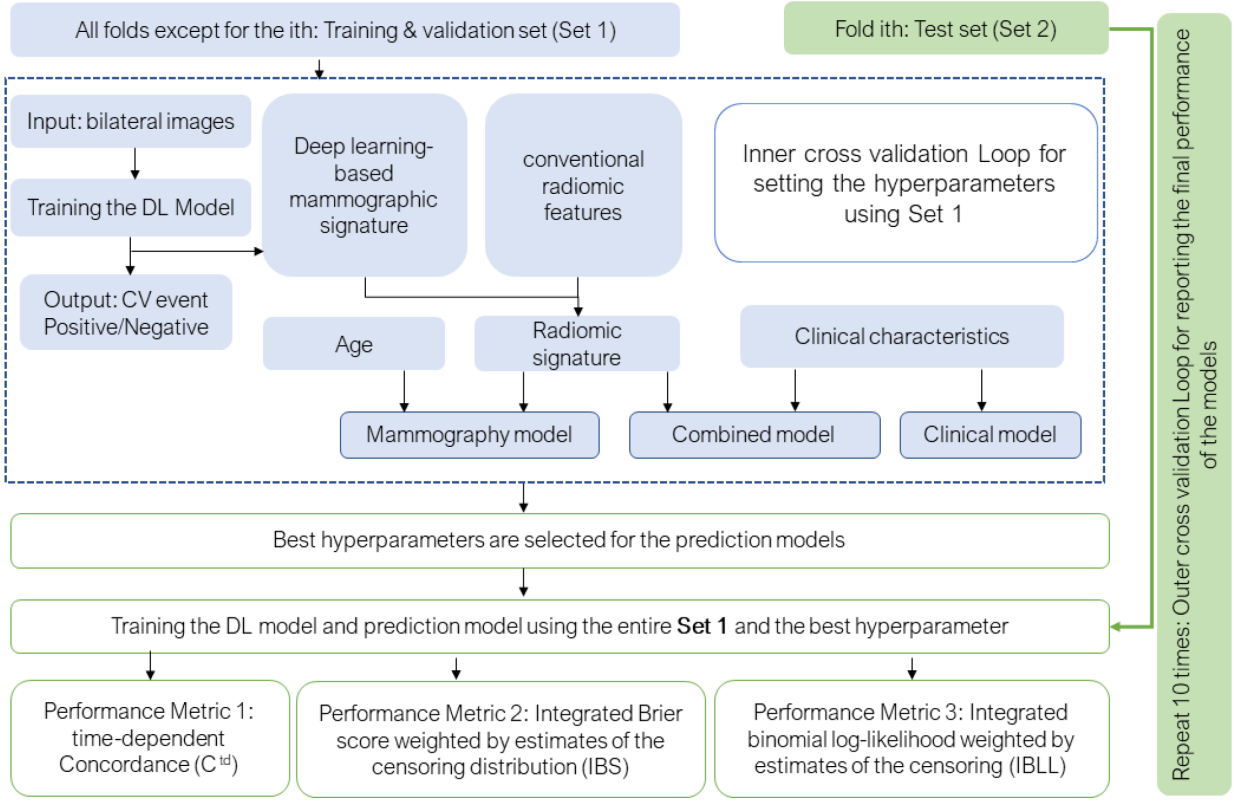


Figure 2 Models and the nested cross-validation process. The mammography model combines radiomic signature with age. For comparison, clinical and combined models were also built. The outer loop is being used to report the final performance metrics of the model, while the inner loop is used for tuning the hyperparameters. The blue and green items refer to the training (inner 10-fold cross-validation loop) and testing (outer cross-validation loop) phase. Both inner and outer loops are repeated 10 times. The same process is done for all three models to ensure the comparability of the models. CV, cardiovascular; DL, deep learning.

Table 1 Participant characteristics and completeness of data

Characteristic	MACE events (n=3392)	No MACE events (n=45 804)	Missing data (n (%)) (n=49 196)
Demographics (mean (SD))			
Age, years	64 (7.5)	60 (6.9)	22 (0.04%)
Height, cm	162 (7.2)	163 (7.2)	2061 (4.2%)
Weight, kg	76 (17.6)	72 (15.3)	1597 (3.3%)
Body mass index, kg/m^2	29 (6.6)	27 (5.7)	2858 (5.8%)
Smoking (%)			354 (0.7%)
Current smoker	278 (8.2%)	2174 (4.7%)	
Ex-smoker	1247 (36.8%)	16 315 (35.6%)	
Never smoker	1833 (54%)	26 995 (58.9%)	
Current drinker	2614 (77.1%)	38 807 (84.7%)	1202 (2.4%)
Medical history (%)			
Breast cancer	36 (1.1%)	178 (0.4%)	1136 (2.3%)
Diabetes	465 (13.7%)	2428 (5.3%)	1126 (2.3%)
Chest radiotherapy	61 (1.8%)	384 (0.8%)	1655 (3.4%)
Breast biopsy	524 (15.4%)	6242 (13.6%)	1736 (3.5%)
Female reproductive history (%)			
Parous	3136 (92.5%)	41 657 (90.9%)	131 (0.27%)
Postmenopausal	2935 (86.5%)	36 866 (80.5%)	1161 (2.4%)
Ever used oral contraceptives	2831 (83.5%)	40 533 (88.5%)	308 (0.6%)
Ever used hormone replacement therapy	1689 (49.8%)	17 862 (39%)	596 (1.2%)
Medication use (%)			
Medications for diabetes	348 (10.3%)	1579 (3.4%)	229 (0.5%)
Medications for high cholesterol	1334 (39.3%)	9711 (21.2%)	229 (0.5%)
Any antihypertensive	1693 (49.9%)	11 712 (25.6%)	229 (0.5%)
Beta blockers	299 (8.8%)	1461 (3.2%)	229 (0.5%)
Calcium channel antagonists	389 (11.5%)	2051 (4.5%)	229 (0.5%)
ACE inhibitors	813 (24%)	5422 (11.8%)	229 (0.5%)
Angiotensin receptor antagonists	879 (25.9%)	6176 (13.5%)	880 (1.8%)
Antiplatelets	1052 (31%)	4455 (9.7%)	229 (0.5%)

Baseline' was defined as time point at which participants consented to participation in the cohort and completed a health questionnaire. 'MACE events' were defined as those who experience an extended MACE during follow-up. 'no MACE' was defined as those who do not experience an extended MACE during follow-up. All variable comparisons between groups yielded significant p<0.0001 except for breast biopsy p=0.001.
MACE, major adverse cardiovascular event.

risk prediction models (using age and clinical data) with their C-index for comparison (table 2). We enabled further comparison with existing standard risk prediction models by using clinical data available for the Lifepool cohort to build a Cox proportional hazards risk prediction model.

RESULTS

Participant characteristics and available data

There were 49 196 women with no evidence of prior cardiovascular disease included in the analysis (table 1). Mean age at baseline was 59.6 (SD 9 years), range 35–94 years and median follow-up was 8.8 years (QR 7.5–10.6 years). At baseline, 5% were current smokers, 62% had body mass index (BMI) >25 kg/m^2, 6% had type 2 diabetes, 33% were taking medication for hypercholesterolaemia, 27% for hypertension and 11% were on an antiplatelet agent. The median time between the first mammogram and the collection of baseline health data was 2 years. There were 3392 individuals that recorded an extended MACE event during follow-up at a rate of 7.6/1000 person years (95% CI 7.4 to 7.9) (table 3). The majority of events were coded as atherosclerotic disease (n=2383) with 731 heart failure events, 656 myocardial infarction events and 434 stroke events also recorded. Prespecified analysis of event rates by age and BMI is shown in online supplemental tables 3 and 4.

Prediction of extended MACE using machine learning

The performance metrics of machine learning models are shown in table 4. The Mammography model built using the age and radiomic data predicted extended MACE events with a concordance index of 0.72 (95% CI 0.71 to 0.73), an integrated Brier score of 0.06 (95% CI 0.058 to 0.063) and an integrated binomial log-likelihood of −0.21 (95% CI −0.22 to −0.20). The slope of the calibration plot was 1.12 (95% CI 1.09 to 1.16). To further visualise the discriminative power of the model, we used the output of the DeepSurv for Mammography model and categorised individuals using the median risk to generate high-risk and low-risk groups (figure 3) (log rank test statistics=76.13, p<0.001). In addition to the primary analyses, we performed subgroup analyses across key subgroups. The concordance index for the Mammography model was 0.72 (95% CI 0.70 to 0.75) among individuals with BMI <25 kg/m^2 and 0.71 (95% CI 0.70 to 0.73) in those with BMI ≥25 kg/m^2. Similarly, the model demonstrated strong performance across menopausal groups, achieving a concordance index of 0.75 (95% CI 0.71 to 0.79) in premenopausal women and 0.71 (95% CI 0.69 to 0.73) in postmenopausal women.

The Clinical model based on clinical characteristics without radiomic data had a concordance index of 0.73 (95% CI 0.72 to 0.74), an integrated Brier score of 0.063 (95% CI 0.059 to

Table 2 Risk prediction model statistics based on each set of features and DeepSurv

	Input features	C^{td}	IBS	IBLL
1	Conventional radiomic signature (two rounds)	0.619 (0.610 to 0.630)	0.062 (0.59 to 0.066)	−0.221 (−0.226 to −0.215)
2	Deep-learnt radiomic signature (two rounds)	0.676 (0.667 to 0.684)	0.064 (0.623 to 0.068)	−0.228 (−0.234 to −0.220)
3	Deep-learnt radiomic signature (only the first round)	0.652 (0.644 to 0.661)	0.066 (0.063 to 0.069)	−0.235 (−0.242 to −0.228)
4	Deep-learnt radiomic signature (only the latest available round)	0.659 (0.651 to 0.668)	0.068 (0.065 to 0.071)	−0.234 (−0.241 to −0.227)
5	Deep-learnt radiomic signature and conventional signature (two rounds)	0.699 (0.692 to 0.708)	0.062 (0.059 to 0.064)	−0.218 (−0.224 to −0.211)
6	Age	0.640 (0.634 to 0.651)	0.066 (0.063 to 0.071)	−0.233 (−0.238 to −0.225)
7	Age and deep-learnt radiomic signature	0.700 (0.692 to 0.709)	0.064 (0.061 to 0.066)	−0.227 (−0.232 to −0.219)
8	Mammography model Age and deep-learnt radiomic signature and conventional signature (two rounds)	0.719 (0.711 to 0.729)	**0.060 (0.058 to 0.063)**	**−0.211 (−0.216 to −0.203)**
9	Clinical model (baseline for comparison)	**0.728 (0.720 to 0.736)**	0.063 (0.059 to 0.065)	−0.224 (−0.212 to −0.230)

For each performance metric, the best-performing value is shown in bold. Comparing rows 2, 3 and 4 suggests that there is a slight but significant improvement when using both screening rounds rather than a single round. Comparing rows 2 and 5 indicates a slight but statistically significant improvement when conventional radiomic features are added to deep learning-based radiomic features. Comparing rows 7 and 8 shows that, even after adding age to the deep learning-based features, incorporating conventional radiomic features results in a further slight but statistically significant improvement.
C^{td}, time-dependent concordance; IBLL, integrated binomial log-likelihood; IBS, integrated Brier score; ICD, International Classification of Diseases.

0.065) and an integrated binomial log-likelihood of −0.22 (95% CI −0.23 to −0.21). The slope of the calibration plot was 1.00 (95% CI 0.95 to 1.063). All performance metrics for the Clinical model are directly comparable to those for the Mammography model. The model built using age alone predicted extended MACE events with a concordance index of 0.64 (95% CI 0.63 to 0.65) (table 2).

The Combined model that used clinical characteristics and radiomic data had a concordance index of 0.75 (95% CI 0.74 to 0.76), an integrated Brier score of 0.058 (95% CI 0.054 to 0.060) and an integrated binomial log-likelihood of −0.21 (95% CI −0.21 to −0.19). The slope of the calibration plot was 0.94 (95% CI 0.87 to 1.00). Both the concordance index and the integrated Brier score were marginally improved with the Combined model, though the integrated binomial log-likelihood and slopes of the calibration plots were broadly similar. Repeating the analyses with the DeepHitSingle model showed comparable results to those obtained with the DeepSurv model (table 2, online supplemental figure 1).

Table 3 Cardiovascular (CV) events and follow-up recorded

	Events (%) among 49 196 participants	Person-years of follow-up	Crude event rate (95% CI) per 1000 person years
Extended MACE	3392 (6.9)	443 603	7.6 (7.4 to 7.9)
Fatal extended MACE	258 (0.5)	456 364	0.5 (0.5 to 0.6)
Non-fatal extended MACE	3292 (6.7)	443 609	7.4 (7.2 to 7.7)
Myocardial infarction	656 (1.3)	453 978	1.4 (1.3 to 1.6)
Stroke	434 (0.9)	455 016	0.95 (0.9 to 1.6)
Atherosclerotic disease	2383 (4.8)	446 874	5.3 (5.1 to 5.5)
Heart failure	731 (1.5)	454 214	1.6 (1.5 to 1.7)

Extended MACE: myocardial infarction (ICD-I21, I22, I23), stroke (ICD-I63-I64), atherosclerotic disease (ICD-I20, I24, I25, I65-I66, I70-I77) and heart failure (ICD-I50) and CV death. CI for crude event rates were calculated using Poisson regression.
MACE, major adverse cardiovascular event.

Prediction of extended MACE using Cox proportional hazards model

The Cox proportional hazards model using baseline participant data had a Harrell's C discrimination index of 0.72 (95% CI 0.72 to 0.73) and a calibration plot's slope of 0.96 (95% CI 0.90 to 1.01). For comparison, the Harrell's C discrimination indices for established risk scores compared with our 'mammographic' model, our 'clinical' risk factor model and our 'combined' model are shown in table 5 and figure 4. Beta coefficients and adjusted HR for the Cox proportional model are shown in online supplemental table 3 and receiver operator curves for all three models are shown in online supplemental figure 2.

DISCUSSION

The primary finding from these analyses was that a deep learning-based model using just routine screening mammography images coupled with age was able to predict the risk of cardiovascular events in women. Further, the deep learning model had measures of accuracy, discrimination and calibration that compared favourably to traditional risk prediction methods including those based on cardiovascular risk factors and Cox proportional hazards models. This was true for comparisons of models based on mammography data and cardiovascular risk factor data made using machine learning within the Lifepool data, as well as for comparisons of machine learning models versus traditional models using the same data. It was also true when comparing the performance of the machine learning-based models derived from the Lifepool data against the performance of traditional risk prediction models generated using a range of external datasets reported in the literature.

There were small improvements in risk prediction metrics achieved with a more complex model that combined both mammography data and traditional cardiovascular risk factor

Table 4 Performance metrics for risk prediction models

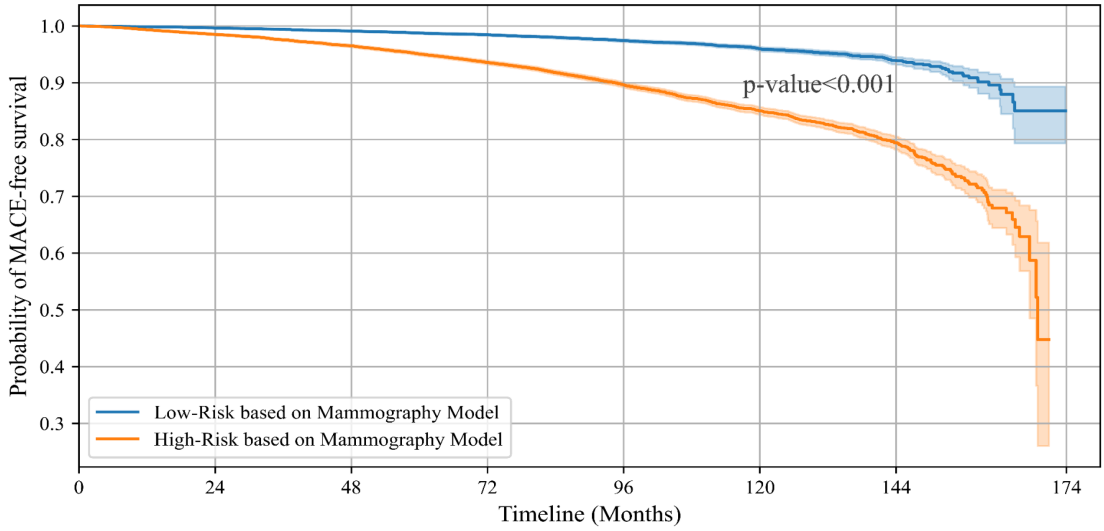
	Time-dependent concordance index (95% CI)	Integrated Brier score (95% CI)	Integrated binomial loglikelihood (95% CI)
Machine learning risk prediction models based on Lifepool data			
Mammography model (age and mammography data)	0.72 (0.71 to 0.73)	0.06 (0.06 to 0.06)	−0.21 (−0.22 to −0.20)
Clinical model (age and clinical characteristics)	0.73 (0.72 to 0.74)	0.06 (0.06 to 0.07)	−0.22 (−0.23 to −0.21)
Combined model (age, mammography data and clinical characteristics)	0.75 (0.74 to 0.76)	0.06 (0.05 to 0.06)	−0.21 (−0.19 to −0.21)

data. There are also models from other jurisdictions with superior metrics to all those achieved in our study, all based on Cox proportional hazards models. The higher performing models use more variables and wider age ranges, resulting in better discrimination,[18–21] but they are typically more resource intensive, less widely used as a consequence and can be inaccurate when available data are incomplete. Our mammographic model with a C-statistic of 0.72 performed well when comparing to models in similar cohorts, for example, the 'PREDICT' prediction model which reports a C-statistic of 0.73 for women.[20]

A key advantage of the mammography model we developed is that it did not require additional history taking or medical record data and leveraged an existing risk screening process widely used by women. Because of its simplicity, a mammography model may have the capacity to serve as a cardiovascular risk prediction tool for women in diverse communities across Australia and around the world. Mammography has potential as a 'two-for-one' risk assessment tool, offering efficiencies for both community and the healthcare system. A future prospective implementation trial with health economic evaluation is recommended to establish the clinical utility, acceptability and cost-effectiveness of mammography-based cardiovascular risk prediction.

Cardiovascular risk prediction is undeused in women for reasons that include gender bias in the delivery of services,[3] lack of resources in the healthcare system and the suboptimal performance of existing risk prediction algorithms in women.[2 22] Efforts to increase cardiovascular screening rates in women have been largely ineffective.[23] A recent WHO report identified global cardiovascular risk screening programmes as inadequate and called for re-evaluation of population-level programmes and consideration of alternative approaches.[24] By comparison, mammography is widely used for breast cancer screening, is often government-funded and has been shown to be highly effective in Australia.[7] Mammography-based screening programmes have engaged women very effectively, with >67% of women in the USA and the UK participating in screening mammography.[7 25 26]

The use of mammography images to predict cardiovascular risk is novel, but the use of machine learning models to do cardiovascular risk prediction is gaining traction. Machine learning models based on cardiovascular risk factors have been shown to outperform traditional risk prediction methods,[19 27 28] with better capacity to model nonlinear relationships between risks and outcomes identified as one possible advantage. Moreover, as seen in the study by Allen *et al*, and our study, using



Low-Risk: At Risk (Censored)	24598 (0)	24391 (114)	23768 (601)	19152 (5067)	11064 (12994)	5403 (18539)	1140 (22735)	0 (23846)
Low-Risk: Events	0	93	229	379	540	656	723	752
High-Risk: At Risk (Censored)	24598 (0)	24019 (179)	22868 (822)	16888 (6166)	8463 (14049)	3663 (18534)	772 (21264)	0 (21958)
High-Risk: Events	0	400	908	1544	2086	2401	2562	2640

Figure 3 Kaplan-Meier plots for the high-risk and low-risk groups based on the output of DeepSurv. Kaplan-Meier plot for the high-risk and low-risk groups. Based on output from DeepSurv which relies on age, deep-learnt and conventional radiomic features and on the median probability as the threshold (log rank test statistics=76.13, p<0.001). MACE, major adverse cardiovascular event.

Table 5 Performance of standard risk prediction models for comparison

Model	C statistic (95% CIs)	Mean age (STD)	Age range (min-max)	Prediction time window
Cox risk prediction model based on lifepool data	0.72 (0.71 to 0.73)	60 (7)	35–94	10 years
Cox risk prediction model based on other data				
SCORE2[6]	0.73 (0.73 to 0.73) and 0.77* (0.76 to 0.77)	Not provided.	40–69	10 years
PCE model[19]	0.70 (0.55 to 0.83)	58 (8)	40–79	10 years
Cho *et al*[28]	0.75 (0.74 to 0.76)	58 (8)	40–79	5 years
PREDICT Model[20]	0.73 (0.72 to 0.73)	56 (9) (women)	30–74	5 years
Q RISK 3[21]	0.86 (0.86 to 0.86) and 0.88* (0.88 to 0.88)	43 (15) (women)	25–84	10 years
WHO CV RISK CHARTS 2019[31]	0.68 (0.63 to 0.74) to 0.83 (0.78 to 0.88)†	56 (9)	48–63	10 years
QR4[6]	0.74 (0.74 to 0.74) 0.78* (0.78 to 0.78)	39 (15)	18–84	10 years
PREVENT[5]	0.76 and 0.79*	53 (13) (women)	30–79	10 years
AusCVDRisk (external validation)	‡	Not available	Not available	5 years

*C statistic for men and women, respectively.
†For different population subsets.
‡There are no external validated datasets for the AusCVDRisk score to date.
CV, cardiovascular.

an automated machine-based methodology for identification of mammographic features for predicting cardiovascular risk is feasible and reliable.[29] Further, by including whole breast mammographic features, our model aims to overcome the limitations of using BAC alone. While BAC is linked with some cardiovascular risk factors, it is negatively associated with others, such as smoking, and has been demonstrated to be less accurate at predicting CV risk in older women.[29] Our model offers a more comprehensive, automated assessment of breast characteristics.

This study benefits from the large sample size, long duration of follow-up, large number of outcome events recorded and the nested cross-validation methodology used for evaluating the model's performance. Nested cross-validation is superior to the traditional 'hold-out' validation method because it maximises data utilisation and reduces variance.[16] The study also has some limitations. Normalisation was used to manage the issue of mammographic images being sourced from multiple different machines. This may be inadequate for ensuring seamless translation of deep learning models across different scanners[30] and risks information loss. Therefore, the deep learnt image features were complemented with conventional radiomic features extracted from the images prior to normalisation. In addition, the cardiovascular risk factors used for the comparator models were self-reported, and thus it was not possible to build machine learning-based models or Cox proportional hazards models for the Lifepool participants that directly matched external cardiovascular risk assessment tools. All deep learning models are restricted by their training datasets, and recalibration may be required for external validation in settings that use different mammographic machines, screening practices or

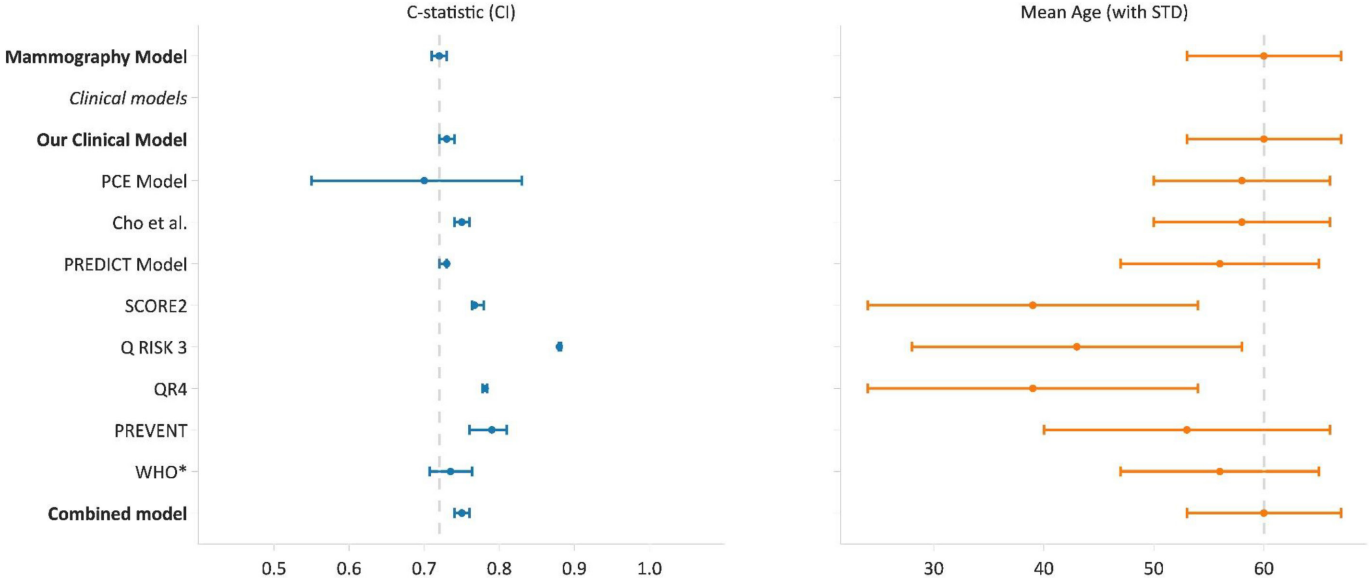


Figure 4 Performance of standard risk prediction models compared with three models developed in this study. The C-statistics and the age ranges are from the references cited in table 2. Where a women-specific model was available, the C-statistic for that model is shown. *The WHO model is based on the model's performance on the APCSC data. The mean age and STD are not provided in the paper, but the age range is 40–69 for this model. 'Mammography model' based on age and radiomic data, 'clinical model' based on clinical characteristics (all variables in table 1) without radiomic data. 'Combined model', based on clinical characteristics and radiomic data for this study.

that assess different ethnic groups. Further, we were not able to directly compare our comprehensive algorithm to one based on BAC alone. Economic evaluation is out of the scope of this study, though a cost–benefit analysis of screening mammography cardiovascular risk prediction models should be performed before clinical application. Moving forward, the algorithm should be externally validated in additional, diverse cohorts, including populations with different ethnic compositions and screening practices, to assess generalisability and the need for recalibration. Implementation research should be undertaken to identify potential barriers and enablers to integration into routine practice.

CONCLUSIONS

A deep learning algorithm utilising routine mammograms and age shows promise as a cardiovascular risk prediction tool. Mammography may offer a cost-effective 'two for one' opportunity to screen women for both breast cancer and cardiovascular risk, enabling broader cardiovascular risk screening for women than is currently achieved.

Author affiliations
[1]Cardiovascular Division, The George Institute for Global Health, Sydney, New South Wales, Australia
[2]Faculty of Medicine and Health, University of New South Wales, Sydney, New South Wales, Australia
[3]Discipline of Clinical Imaging, Faculty of Medicine and Health, The University of Sydney, Sydney, New South Wales, Australia
[4]University of New South Wales, Sydney, New South Wales, Australia
[5]Queensland Digital Health Centre, University of Queensland, Brisbane, Queensland, Australia
[6]Professorial Unit, The George Institute for Global Health, Sydney, New South Wales, Australia
[7]School of Medicine and Psychology, Australian National University, Canberra, Australian Capital Territory, Australia
[8]Faculty of Medicine and Health Sciences, The University of Auckland, Auckland, New Zealand
[9]The George Institute for Global Health, Sydney, New South Wales, Australia
[10]The George Institute for Global Health, School of Public Health, Imperial College London, London, UK
[11]School of Public Health, Imperial College London, London, UK

Acknowledgements All authors meet criteria for authorship and had full access to the data in the study and accept responsibility for the submission of this article. We would like to acknowledge the contribution of PETER MAC and Lisa Devereux, Professor John Hopper and the Data Management team.

Contributors All designated authors have met the four ICMJE criteria for authorship including substantial contributions to the conception or design of the work; or the acquisition, analysis or interpretation of data for the work; drafting the work or reviewing it critically for important intellectual content; final approval of the version to be published; agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. JYB and ZG stand as guarantors for the above. PB and CA are joint last authors.

Funding The funding for this project arises from an MRFF Grant No. 1201433. ZG is supported by the New South Wales Cancer Institute Early Career Fellowship scheme (2022/ECF1426). MW is supported by National Health and Medical Research Council grant APP1174120. JYB is supported by a New South Wales Health Early-Mid Career Research Grant (2022/RG212206). KKP is supported in part by a Heart Foundation Heart Health Research Trust senior fellowship (1886).

Disclaimer The funding bodies had no role in study design; collection, analysis or interpretation of data; or writing of the manuscript.

Competing interests JYB does not have any conflicts of interest to disclose. She is supported by a New South Wales Health Early-Mid Career Research Grant (2022/RG212206). ZG is chief scientific officer of DetectedX, a cloud-based educational company. She is supported by the New South Wales Cancer Institute Early Career Fellowship scheme (2022/ECF1426). RAF does not have any conflicts of interest to disclose. SB does not have any conflicts of interest to disclose. NI-HK does not have any conflicts of interest to disclose. AR does not have any conflicts of interest to disclose. KD does not have any conflicts of interest to disclose. KKP does not have any conflicts of interest to disclose. MW has previously been a consultant to Amgen and Freeline. He is supported by National Health and Medical Research Council grants APP1149987 and APP1174120. BGL does not have any conflicts of interest to disclose. BN does not have any conflicts of interest to disclose. LJ does not have any conflicts of interest to disclose. PB is CEO of DetectedX, a cloud-based educational company. CA has received honoraria or has served on Steering Committees/Advisory Boards for Astra Zeneca, Novo Nordisk and Amgen. She is an employee of the George Institute for Global Health and holds an MRFF Investigator Grant and NSW Health Grant.

Patient and public involvement Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

Patient consent for publication Not applicable.

Ethics approval Ethics approval was obtained from the Peter MacCallum (LP036; HREC/72045/PMCC-2021) and University of Sydney (Project No. 2020/791) ethics committees. All participants provided prior written informed consent.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. Access to mammographic examinations and participants' responses to the survey questions can be requested from the data custodians of the project.

ORCID iDs
Jennifer Yvonne Barraclough http://orcid.org/0000-0002-5093-9900
Sebastiano Barbieri http://orcid.org/0000-0002-5919-372X
Nicholas I-Hsien Kuo http://orcid.org/0000-0001-8749-7280
Katrina K Poppe http://orcid.org/0000-0002-4418-4476
Mark Woodward http://orcid.org/0000-0001-9800-5296
Louisa Jorm http://orcid.org/0000-0003-0390-661X

REFERENCES

1 Vaduganathan M, Mensah GA, Turco JV, *et al*. The Global Burden of Cardiovascular Diseases and Risk: A Compass for Future Health. *J Am Coll Cardiol* 2022;80:2361–71.
2 Collins GS, Altman DG. An independent external validation and evaluation of QRISK cardiovascular risk prediction: a prospective open cohort study. *BMJ* 2009;339:b2584.
3 Hyun KK, Redfern J, Patel A, *et al*. Gender inequalities in cardiovascular risk factor assessment and management in primary healthcare. *Heart* 2017;103:492–8.
4 Australian Institute of Health and Welfare. Heart, stroke and vascular disease—Australian facts. Canberra AIHW; 2021.
5 Khan SS, Matsushita K, Sang Y, *et al*. Development and Validation of the American Heart Association's PREVENT Equations. *Circulation* 2024;149:430–49.
6 Hippisley-Cox J, Coupland CAC, Bafadhel M, *et al*. Development and validation of a new algorithm for improved cardiovascular risk prediction. *Nat Med* 2024;30:1440–7.
7 Australian Institute of Health and Welfare. BreastScreen australia monitoring report. Canberra AIHW; 2022.
8 Ibrahim M, Suleiman ME, Gandomkar Z, *et al*. Associations of Breast Arterial Calcifications with Cardiovascular Disease. *J Womens Health (Larchmt)* 2023;32:529–45.
9 Lee SC, Phillips M, Bellinge J, *et al*. Is breast arterial calcification associated with coronary artery disease?-A systematic review and meta-analysis. *PLoS One* 2020;15:e0236598.
10 Grassmann F, Yang H, Eriksson M, *et al*. Mammographic features are associated with cardiometabolic disease risk and mortality. *Eur Heart J* 2021;42:3361–70.
11 Ahn S, Chang Y, Kwon R, *et al*. Mammography-based deep learning model for coronary artery calcification. *Eur Heart J Cardiovasc Imaging* 2024;25:456–66.
12 Wang K, Khan N, Highnam R. Automated segmentation of breast arterial calcifications from digital mammography. 2019 International Conference on Image and Vision Computing New Zealand (IVCNZ); Dunedin, New Zealand. 10.1109/IVCNZ48456.2019.8960956 Available: https://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=8955481
13 Katzman JL, Shaham U, Cloninger A, *et al*. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Med Res Methodol* 2018;18:24.
14 Kvamme H, Borgan Ø, Scheel I. Time-to-Event Prediction with Neural Networks and Cox Regression. *J Mach Learn Res* 2019;20:1–30.

15 Kvamme H, Borgan Ø. Continuous and discrete-time survival prediction with neural networks. *Lifetime Data Anal* 2021;27:710–36.
16 Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics* 2006;7:1–8.
17 Lee C, Zame W, Yoon J, *et al*. DeepHit: A Deep Learning Approach to Survival Analysis With Competing Risks. *AAAI* 2018;32.
18 Hageman S, Pennells L, Ojeda F, *et al*. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. *Eur Heart J* 2021;42:2439–54.
19 Goff DC Jr, Lloyd-Jones DM, Bennett G, *et al*. 2013 ACC/AHA Guideline on the Assessment of Cardiovascular Risk. *Circulation* 2014;129:S49–73.
20 Pylypchuk R, Wells S, Kerr A, *et al*. Cardiovascular disease risk prediction equations in 400 000 primary care patients in New Zealand: a derivation and validation study. *Lancet* 2018;391:1897–907.
21 Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ* 2017;357:j2099.
22 Hespe CM, Giskes K, Harris MF, *et al*. Findings and lessons learnt implementing a cardiovascular disease quality improvement program in Australian primary care: a mixed method evaluation. *BMC Health Serv Res* 2022;22:108.
23 Vogel B, Acevedo M, Appelman Y, *et al*. The Lancet women and cardiovascular disease Commission: reducing the global burden by 2030. *Lancet* 2021;397:2385–438.
24 Eriksen CU, Rotar O, Toft U, *et al*. What is the effectiveness of systematic population-level screening programmes for reducing the burden of cardiovascular diseases? 2021. Available: https://apps.who.int/iris/handle/10665/338530
25 Centers for Disease Control and Prevention. Health, United States, 2020-2021. National Center for Health Statistics; 2021. Available: https://www.cdc.gov/nchs/hus/topics/mammography.htm#:~:text=The%20age%2Dadjusted%20percentage%20of,within%20the%20past%202%20years
26 Schünemann HJ, Lerda D, Quinn C, *et al*. Breast Cancer Screening and Diagnosis: A Synopsis of the European Breast Guidelines. *Ann Intern Med* 2020;172:46–56.
27 Barbieri S, Mehta S, Wu B, *et al*. Predicting cardiovascular risk from national administrative databases using a combined survival analysis and deep learning approach. *Int J Epidemiol* 2022;51:931–44.
28 Cho S-Y, Kim S-H, Kang S-H, *et al*. Pre-existing and machine learning-based models for cardiovascular risk prediction. *Sci Rep* 2021;11:8886.
29 Allen TS, Bui QM, Petersen GM, *et al*. Automated Breast Arterial Calcification Score Is Associated With Cardiovascular Outcomes and Mortality. *JACC Adv* 2024;3:101283.
30 Jiang Z, Gandomkar Z, Trieu PDY, *et al*. Evaluating Recalibrating AI Models for Breast Cancer Diagnosis in a New Context: Insights from Transfer Learning, Image Enhancement and High-Quality Training Data Integration. *Cancers (Basel)* 2024;16:322.
31 Kaptoge S, Pennells L, Bacquer D. World Health Organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions. *Lancet Glob Health* 2019;7:e1332–45.