

Deep Learning Analysis of Breast Arterial Calcification to Enhance Cardiovascular Risk Prediction in Women

Mu'ath Ibrahim^{1,2} · Patrick C. Brennan^{1,2} · Mo'ayyad E. Suleiman^{1,2} · Mary Rickard^{1,2} · Clare Arnott³ · Jennifer Y. Barraclough³ · Seyedamir Tavakoli Tabai¹ · Ziba Gandomkar^{1,2}

¹ The University of Sydney, Sydney, Australia

² DetectedX, Sydney, Australia

³ The George Institute for Global Health, Sydney, Australia

PURPOSE

Cardiovascular disease (CVD) is the leading cause of mortality among women worldwide, exceeding the combined burden of all cancers. A substantial proportion of cardiovascular events in women occur in the absence of conventional risk factors, and existing risk-prediction algorithms have repeatedly been shown to underestimate risk in this population. There is therefore a clear unmet need for **sex-specific, opportunistic risk markers** that can be obtained from routinely-acquired imaging.

Breast arterial calcification (BAC) is an incidental finding on mammography, arising from medial calcification of the breast arteries, and has been linked to atherosclerotic cardiovascular disease across multiple prior cohorts. In current practice, however, BAC is reported as a binary present/absent observation — a categorisation prone to inter-reader variability and one that discards information about the *extent and density* of calcification.

We hypothesised that **automated, deep-learning-based segmentation and severity grading of BAC** would capture additional risk information beyond binary presence, and that the resulting grades would **independently predict major adverse cardiovascular events (MACE)** after adjustment for established cardiovascular risk factors.

METHODS — STAGE 1: TRAINING

Building the segmentation model. The **SegModel dataset** comprised 1,270 BAC-positive mammographic images drawn from the Lifepool / BreastScreen Australia archive. Calcified regions were manually delineated using Label Studio by a medical imaging scientist, and the resulting annotations were reviewed by a senior radiologist with over thirty years of clinical experience to confirm anatomical and pathological plausibility.

Critically, the dataset was assembled to span **eleven distinct mammography unit types from seven manufacturers** — Agfa, Fuji, Hologic, Konica Minolta, Philips, Sectra and Siemens. This deliberate vendor heterogeneity addresses a key limitation of earlier BAC-segmentation studies, most of which trained and validated on single-vendor data and demonstrated correspondingly limited cross-vendor generalisability.

A **U-Net** architecture — four-level encoder–decoder with skip connections — was implemented in PyTorch and trained at 1024 × 1024 resolution. The objective function combined binary cross-entropy with the Jaccard loss to balance pixel-wise accuracy against overlap with the sparse ground-truth structure. Optimisation used Adam (learning rate 1×10^{-5}). To preserve anatomical orientation in standard CC and MLO views, augmentation was restricted to intensity-based transformations (HSV brightness, saturation, hue). The dataset was partitioned 70 % / 20 % / 10 % into training, validation, and test subsets at the participant level (see **Figure 1, Stage 1**).

METHODS — STAGE 2: PREDICTION

Linking segmented BAC to cardiovascular outcomes. The trained U-Net was applied — without further fine-tuning — to a separate cohort, **MACEPred**, comprising 9,648 women: 3,190 cases with linked extended MACE outcomes and 6,458 controls matched on age and mammography-unit type (case-to-control ratio approximately 1:2). Mean age was 63.7 ± 7.8 years; mean follow-up was 6.09 years. The two datasets shared no participants, eliminating leakage between model development and outcome analysis.

For each examination we computed two quantitative descriptors: **BAC area**, expressed as the proportion of total breast area occupied by segmented calcification, and **mean BAC intensity**, computed from raw pixel values within segmented regions and z-score–normalised per mammography unit type to neutralise vendor-specific signal characteristics. These features were translated into four grading schemes: (i) *binary present/absent*; (ii) *area-based four-level grading (absent / mild / moderate / severe)*; (iii) *intensity-based four-level grading using tertile-defined cut-points*; and (iv) *a combined scheme in which the area grade is up-rated when intensity is higher*.

Four **Cox proportional hazards models** — one per grading scheme — were fitted to estimate adjusted hazard ratios for MACE, adjusting for age, BMI, smoking status, alcohol intake, and dispensed cardiovascular medications (antiplatelets, beta-blockers, calcium channel blockers, ACE inhibitors, angiotensin II receptor blockers, and therapies for diabetes and hypercholesterolaemia) used as pragmatic proxies for treated cardiovascular risk factors. The composite endpoint encompassed ICD-10–coded ischaemic, cerebrovascular and peripheral arterial disease together with cardiovascular mortality from the National Death Index (**Figure 1, Stage 2**).

RESULTS

Segmentation performance. On the held-out test set the U-Net achieved a **Jaccard similarity coefficient of 0.582, F1 score of 0.756, precision of 0.801, recall of 0.716, and overall accuracy of 0.996**. These overlap-based metrics are comparable to, and in several respects exceed, those reported for the previously-published SCU-Net model (Guo et al., Jaccard 0.581, F1 0.729), despite our use of a substantially more vendor-diverse training cohort. Representative segmentation outputs across all four severity grades are shown in **Figure 2**.

Risk prediction. The prevalence of BAC in MACEPred was **26.96 %**, in line with previously-reported figures for women in this age range. After multivariable adjustment, BAC presence (binary) was independently associated with MACE — **adjusted HR 1.16 (95% CI 1.10–1.23)**.

Crucially, a clear dose–response relationship was observed across severity levels under all multi-class grading schemes. For area-based grading, adjusted hazard ratios scaled progressively from **1.13** (mild) through **1.30** (moderate) to **1.58** (severe). Intensity-based grading showed a smaller but consistent gradient (1.08 → 1.10 → 1.18). The **combined area + intensity scheme produced the steepest gradient (1.10 → 1.25 → 1.65)**, with point estimates progressing monotonically and confidence intervals consistent with a non-linear increase in risk at the severe end (**Figure 3**).

CONCLUSIONS

This study presents a **fully-automated deep-learning framework** for detection, segmentation and severity grading of breast arterial calcification on screening mammograms, and demonstrates that the resulting grades carry independent cardiovascular prognostic information in a multi-vendor population-based cohort. Combined area- and intensity-based grading reflected progressively increasing MACE risk across severity strata, with hazard ratios consistent with the non-linear dose–response relationship reported for coronary calcium scoring.

Because BAC grading can be derived **opportunistically from routine mammograms** — without additional imaging, contrast, or radiation exposure — the framework is well-positioned for integration into existing screening workflows, where it could operate as a background PACS analysis returning structured cardiovascular risk information at no incremental procedural cost. The proposed approach is therefore a strong candidate **opportunistic risk enhancer** for women, particularly within the asymptomatic population where conventional risk equations have repeatedly underperformed.

FUTURE WORK

Direct clinical risk-factor measurements. Current adjustment relied on dispensed-medication records as pragmatic proxies for treated hypertension, dyslipidaemia and diabetes. Linkage to electronic health records will permit incorporation of measured blood pressure, lipid profiles and HbA1c at no additional participant burden, allowing direct quantification of the incremental risk-reclassification value of BAC grading over established clinical scores.

Longitudinal BAC progression. The present analysis was based on a single screening round and therefore cannot speak to the prognostic value of **BAC progression** between serial mammograms. A planned longitudinal extension will examine whether change in BAC grade across screening rounds carries independent risk information beyond a single time-point assessment.

Radiomic extension and prospective validation. Future work will move beyond area and intensity to incorporate **radiomic descriptors** (shape, branching pattern, vessel-length distribution), and will validate the framework prospectively in additional international cohorts. A planned multi-site rollout on the **DetectedX** platform will allow BAC grading to be delivered alongside routine breast-cancer screening as a structured cardiovascular flag within the radiologist's report.

Figure 1. Two-stage deep-learning pipeline

Stage 1 trains the U-Net on hand-annotated mammograms. Stage 2 applies it to an unseen cohort and turns the segmented calcium into an adjusted hazard ratio for MACE.

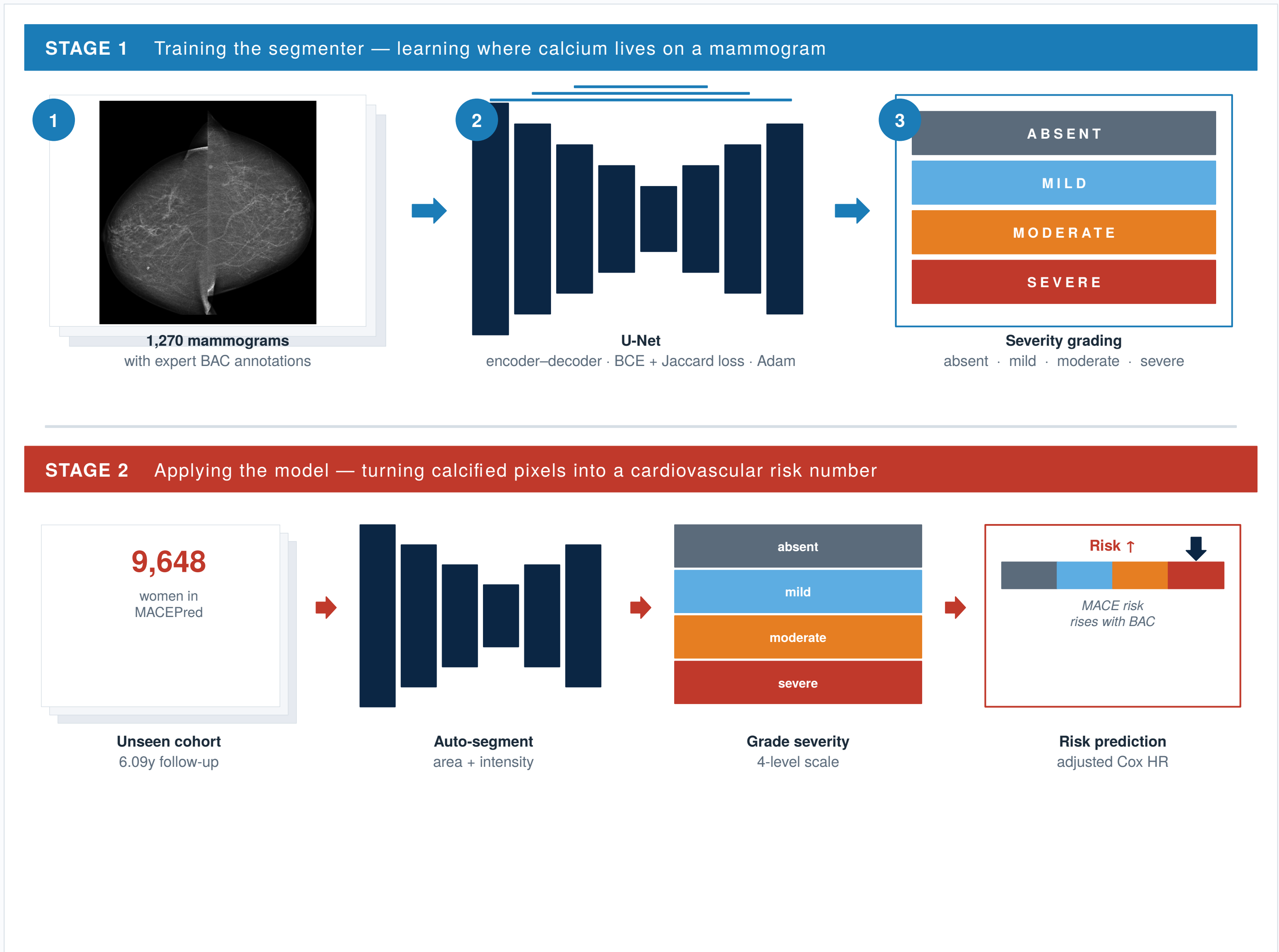


Figure 2. Representative segmentation outputs

Four BAC severity grades. Per case, panels show original mammogram, expert ground-truth annotation, and U-Net–predicted segmentation.

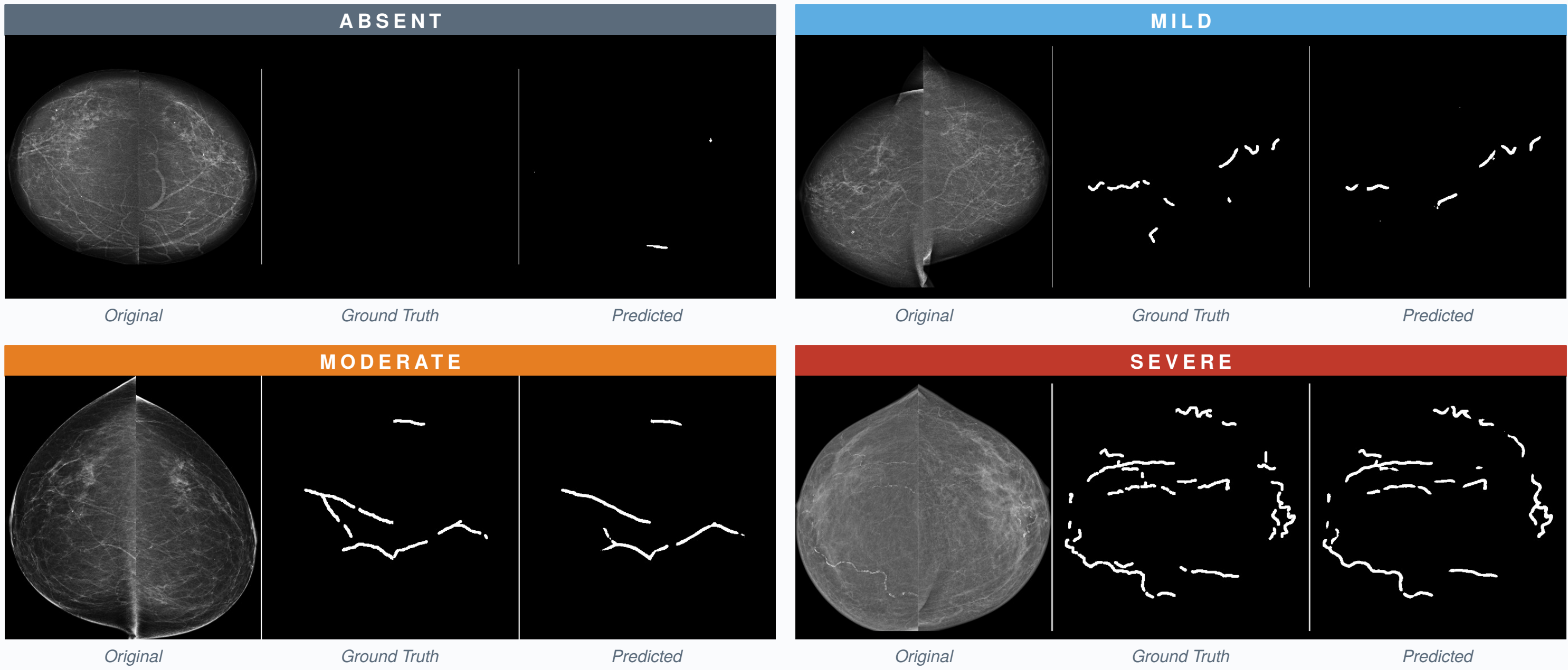


Figure 3. Adjusted hazard ratios by combined BAC grade

Combined area + intensity grading scheme — Cox model adjusted for age, BMI, smoking, alcohol, and medications. BAC absent serves as the reference.

