Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Histopathological Spectrum of Breast Lesions and Factors Associated with Malignancy in Mogadishu, Somalia: A Retrospective Cross-Sectional Study

Mohamed Serar Yusuf

indhooska@simad.edu.so

SIMAD University

Mohamed Husein Adam

SIMAD University

Rahma Mohamed Ali

SIMAD University

Abdkerem Abdulahi Abdi

SIMAD University

Research Article

Keywords: breast cancer, fibroadenoma, histopathology, logistic regression, Somalia, sub-Saharan Africa, low-resource settings

Posted Date: July 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10220161/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Breast pathology in Somalia remains poorly characterized, with few studies examining the histopathological spectrum or factors associated with malignancy in this post-conflict, low-resource setting. This study described the distribution of breast lesion diagnoses and identified factors independently associated with malignant outcomes in Mogadishu, Somalia.

Methods

A retrospective cross-sectional study utilized secondary histopathological data from 522 breast specimens collected across multiple healthcare facilities in Mogadishu (July 2021–December 2025). Multivariable binary logistic regression was applied following systematic data cleaning and bivariate analyses. Model discrimination was assessed by AUC with 95% bootstrap CI, Nagelkerke R^2, and the Hosmer–Lemeshow (HL) test. Four pre-specified sensitivity analyses were conducted.

Results

Of 522 cases (97.1% female; median age 35 years, IQR: 24–50), 437 had definitive outcomes: 262 benign (59.9%) and 175 malignant (40.1%). Fibroadenoma/adenoma was the most common diagnosis (n = 173, 33.1%), followed by invasive ductal carcinoma/NST (n = 101, 19.3%). Malignant cases were significantly older (median 45 vs. 25 years; $p < 0.001$) and presented with larger tumors (median 5.0 vs. 3.5 cm; $p < 0.001$). In multivariable analysis (n = 369 complete cases), increasing age (OR 1.11, 95% CI: 1.08–1.13, $p < 0.001$) and tumor size > 5 cm (OR 1.90, 95% CI: 1.06–3.39, $p = 0.030$) were independently associated with malignancy, while tumor size ≤ 2 cm was inversely associated (OR 0.15, 95% CI: 0.06–0.41, $p < 0.001$). The model demonstrated excellent discrimination (AUC = 0.876, 95% CI: 0.841–0.909; Nagelkerke R^2=0.493). The HL test indicated calibration mismatch (χ^2=22.6, $p = 0.004$), consistent with the parsimonious exploratory model. Sensitivity analyses confirmed robustness across all scenarios.

Conclusion

Fibroadenoma dominates among younger women, while malignancy — predominantly invasive ductal carcinoma — presents with larger tumor size among older patients. Age and tumor size were independently associated with malignant outcome. These findings support heightened awareness and earlier clinical evaluation among women aged approximately 35 years and older, alongside investment in standardized histopathological reporting infrastructure in Somalia and comparable settings.

1. INTRODUCTION

Breast cancer is the most commonly diagnosed cancer among women globally, accounting for approximately 2.3 million new cases and 685,000 deaths in 2020 [1]. While incidence rates are highest in high-income countries, mortality disproportionately burdens low- and middle-income countries (LMICs), largely due to late-stage diagnosis, limited pathological services, and fragile health systems [2,3]. In sub-Saharan Africa, over 75% of breast cancer cases are diagnosed at advanced stages, a pattern strongly linked to poor survival outcomes [2].

Somalia presents a particularly challenging context. Decades of civil conflict, displacement, and infrastructure collapse have severely constrained health data collection, diagnostic capacity, and cancer registry development [4]. Few studies have comprehensively examined the histopathological spectrum of breast disease or identified clinical factors associated with malignant outcomes in this setting. Understanding the distribution of benign versus malignant breast lesions and the patient-level characteristics associated with malignancy is fundamental to designing effective early detection and referral strategies appropriate to local epidemiology.

Secondary histopathological datasets from pathology laboratories offer a practical avenue to generate evidence where prospective data collection remains operationally challenging. This study leveraged a secondary dataset spanning five years (2021–2025) from multiple healthcare facilities in Mogadishu to: (i) describe the sociodemographic and clinical characteristics of patients presenting with breast lesions; (ii) characterize the histopathological spectrum of breast diagnoses; and (iii) identify factors independently associated with malignant outcomes using multivariable logistic regression with rigorous sensitivity analyses.

Regionally, the burden of breast cancer in Africa has become increasingly visible through improved registry reporting, GLOBOCAN estimates, and multicountry studies. Contemporary estimates identify breast cancer as the leading cancer among women in Somalia and across much of Eastern Africa, but mortality remains high because many patients enter care after a clinically detectable mass has already become large or locally advanced [11–14]. In this context, facility-based pathology datasets are valuable because they capture the real diagnostic workload faced by laboratories and provide a practical foundation for local cancer-control planning where population-based cancer registries are limited [15–18].

Evidence from East, West, Central, and Southern Africa consistently shows two patterns that are directly relevant to the present study: benign breast disease is common among adolescents and young women, with fibroadenoma usually the leading diagnosis, while malignant cases are dominated by invasive ductal carcinoma/no special type and frequently present at a younger age and with larger tumor size than is typical in high-income settings [19–27]. Similar observations from Uganda, Kenya, Ethiopia, Rwanda, Tanzania, Nigeria, Ghana, Sudan, and other African settings highlight the need for locally grounded histopathological data rather than relying only on estimates from high-income countries [28–36].

Somalia-specific evidence remains limited but is growing. Available reports describe major health-system barriers, incomplete cancer surveillance, low screening capacity, and shortages of immunohistochemistry and oncology infrastructure [15–18,37–39]. These constraints make routine variables such as age, tumor size, laterality, diagnosis category, and metastasis notation particularly important for early local research. Therefore, expanding the evidence base from Mogadishu can support context-appropriate referral pathways, strengthen pathology reporting standards, and identify priority areas for future prospective studies that include receptor testing, stage, treatment, and survival outcomes [40–45].

2. METHODS

2.1 Study Design and Setting

This retrospective cross-sectional study utilized secondary histopathological data from breast specimens collected across Nova Diagnostic Center in Mogadishu, Somalia, between July 2021 and December 2025. Histopathological examinations were performed at Nova Diagnostic Center, a private pathology laboratory in Mogadishu that receives specimens from multiple public and private healthcare facilities throughout the city. Accordingly, although pathology reporting was centralized at a single laboratory, the study population represents patients referred from multiple healthcare institutions.

Data were extracted from institutional pathology records. Full details of ethical approval and data governance are provided in the Ethical Considerations section.

2.2 Data Source and Participants

The dataset comprised 522 consecutive breast pathology cases with documented histopathological diagnoses. Cases with concomitant non-breast primary pathologies (e.g., thyroid carcinoma identified incidentally during breast workup) were retained with appropriate notation. All sexes presenting with breast pathology were eligible for inclusion.

2.3 Data Cleaning and Variable Harmonization

Systematic data cleaning was performed in Python (version 3.12), including: (i) conversion of a non-numeric age entry to missing; (ii) standardization of hospital names to resolve typographical duplicates; (iii) harmonization of laterality (14 raw variants collapsed into Left, Right, Bilateral, Other/Unknown), lymphovascular invasion (LVI), and perineural invasion (PNI); (iv) standardization of tumor grade into a six-level taxonomy; and (v) reclassification of 310 unique free-text histopathological diagnoses into 12 clinically meaningful categories. Histopathological diagnoses were independently reviewed by the investigators using WHO Classification of Breast Tumors (5th edition) and grouped into clinically meaningful diagnostic categories before analysis. Tumor size was categorized as ≤ 2 cm (T1), 2–5 cm (T2), and > 5 cm (T3/T4). The primary binary outcome was: Malignant = 1 (inclusive of ‘Malignant’ and ‘Metastatic’ coded outcomes) and Benign = 0. Cases with indeterminate outcomes (n = 59) and missing

outcome data (n = 26) were excluded from the primary analysis but included in a pre-specified sensitivity analysis.

2.4 Statistical Analysis

Continuous variables were summarized as medians with interquartile ranges (IQRs). Categorical variables were reported as frequencies and proportions. Mann–Whitney U tests and chi-squared tests (with Yates' continuity correction; Fisher's exact test where expected cell counts < 5) were used for bivariate comparisons. Multivariable binary logistic regression was used to identify factors independently associated with malignant outcome. Age (continuous) and tumor size category (reference: T2, 2–5 cm) were selected a priori based on clinical plausibility and data completeness; metastasis status was excluded due to quasi-complete separation with the outcome. Model discrimination was assessed by AUC with 95% CI derived from 2,000 bootstrap resamples; calibration by the HL test (10 decile groups); and effect size by Nagelkerke R^2. Given the exploratory objective, the model was used to characterize associations rather than derive individual risk scores. Complete-case restriction and the resulting analytic sample size are reported transparently. Statistical significance was set at $\alpha = 0.05$; all analyses were conducted in Python using scipy, statsmodels, and scikit-learn.

Four pre-specified sensitivity analyses were conducted: (SA1) restriction to female patients only; (SA2) exclusion of extreme tumor size outliers (> 20 cm); (SA3) restriction to adult patients (age ≥ 18 years); and (SA4) use of a broadened outcome definition, reclassifying Suspicious, Atypical, and Atypia outcomes as malignant to evaluate the potential impact of outcome misclassification.

Variables with near-universal missingness — Residence (100%), Insurance_Status (100%), ER/PR/HER2 status (> 99%), and Ki67 (> 99%) — were excluded from multivariable modeling and are documented as study limitations. The complete-case analytic sample and the number of observations excluded relative to the definitive-outcome dataset are reported explicitly.

3. RESULTS

3.1 Study Population

A total of 522 breast pathology cases were included from July 2021 to December 2025 (Table 1). The vast majority were female (n = 507, 97.1%), with 15 male cases (2.9%). Median age was 35 years (IQR: 24–50; range: 8–92); three cases had missing age data. The highest frequency was in the 20–29 year age band (n = 133, 25.5%), followed by 30–39 years (n = 111, 21.3%); 63 cases (12.1%) were under 20 years. Among 496 cases with documented facility type, the vast majority attended tertiary facilities (n = 481, 97.0%).

Table 1
Sociodemographic and Clinical Characteristics of the Study Population (N = 522)

Characteristic	Category / Statistic	n (%)
Sex	Female	507 (97.1%)
	Male	15 (2.9%)
Age (years)	Median (IQR)	35 (24–50)
	Range	8–92
	Missing	3 (0.6%)
Age Group	< 20 years	63 (12.1%)
	20–29 years	133 (25.5%)
	30–39 years	111 (21.3%)
	40–49 years	81 (15.5%)
	50–59 years	56 (10.7%)
	60–69 years	39 (7.5%)
	≥ 70 years	36 (6.9%)
Study Period	July 2021 – December 2025	—
Healthcare Facility†	Tertiary	481 (97.0%)
	Secondary	8 (1.6%)
	Specialized	6 (1.2%)
	Clinic	1 (0.2%)
	Missing	26 (5.0% of N = 522)
Laterality‡	Left	232 (46.1%)
	Right	209 (41.6%)
	Bilateral	30 (6.0%)
	Other/Unknown	32 (6.4%)
	Missing	19 (3.6% of N = 522)
Tumor Size (cm)§	Median (IQR)	4.0 (2.5–6.0)

†Percentages for facility type among 496 cases with documented facility. ‡Percentages among 503 cases with documented laterality. §Tumor size percentages among 422 cases with recorded size. *Indeterminate: Suspicious (n = 33), Atypical (n = 11), Atypia (n = 4), Inadequate (n = 5), Borderline (n = 3), Non-diagnostic (n = 2), Inconclusive (n = 1).

Characteristic	Category / Statistic	n (%)
	≤ 2 cm (T1)	70 (16.6%)
	2–5 cm (T2)	219 (51.9%)
	> 5 cm (T3/T4)	133 (31.5%)
	Missing	100 (19.2% of N = 522)
Outcome	Benign	262 (59.9% of N = 437)
	Malignant	175 (40.1% of N = 437)
	Indeterminate (excl. from primary model)*	59 (11.3% of N = 522)
	Missing	26 (5.0% of N = 522)

†Percentages for facility type among 496 cases with documented facility. ‡Percentages among 503 cases with documented laterality. §Tumor size percentages among 422 cases with recorded size. *Indeterminate: Suspicious (n = 33), Atypical (n = 11), Atypia (n = 4), Inadequate (n = 5), Borderline (n = 3), Non-diagnostic (n = 2), Inconclusive (n = 1).

3.2 Histopathological Spectrum

The most common category was fibroadenoma and related adenomas (n = 173, 33.1%; Fig. 1). Invasive ductal carcinoma/NST accounted for 101 cases (19.3%), followed by other benign lesions (n = 64, 12.3%), ductal carcinoma (n = 50, 9.6%), and cases suspicious for malignancy (n = 20, 3.8%). Metastatic carcinoma and phyllodes tumors each represented 14 cases (2.7%). Rarer subtypes included papillary carcinoma (n = 6), invasive lobular carcinoma (n = 5), and mucinous carcinoma (n = 2).

3.3 Clinical and Pathological Characteristics

Tumor size was recorded in 422 cases (80.8%); median 4.0 cm (IQR: 2.5–6.0; range: 0.1–60 cm). T2 (2–5 cm) was most frequent (n = 219, 51.9%), followed by T3/T4 (n = 133, 31.5%) and T1 (n = 70, 16.6%). Of 503 cases with documented laterality, 232 were left-sided (46.1%), 209 right-sided (41.6%), 30 bilateral (6.0%), and 32 other/unspecified (6.4%). LVI was documented in 67 cases: 35 positive (52.2%) and 32 negative (47.8%). PNI was available for 36 cases: 13 positive (36.1%), 23 negative (63.9%). IHC receptor status (ER, PR, HER2) and Ki67 were available in fewer than five cases each and could not contribute to inferential analyses (see Limitations). Of 500 cases with recorded metastasis status, 45 (9.0%) had confirmed metastasis, 22 (4.4%) possible/uncertain nodal involvement, and 433 (86.6%) no metastasis. Pathological staging was available in only 65 cases (12.5%).

3.4 Outcome Classification and Exclusions

Of 496 cases with recorded outcomes, 261 were classified benign and one as ‘Benign breast; thyroid malignancy’ (combined: benign group, n = 262), while 167 were malignant and eight metastatic (combined: malignant group, n = 175). For the primary binary analysis, 437 cases had definitive outcomes: 262 benign (59.9%) and 175 malignant (40.1%). A further 59 cases with indeterminate

outcomes (Suspicious n = 33, Atypical n = 11, Atypia n = 4, Inadequate n = 5, Borderline n = 3, Non-diagnostic n = 2, Inconclusive n = 1) and 26 cases with missing outcomes were excluded from the primary model; sensitivity analysis SA4 evaluated the impact of reclassifying indeterminate cases.

3.5 Bivariate Comparisons

Malignant cases were significantly older than benign cases (median 45 years [IQR: 36–57] vs. 25 years [IQR: 20–36]; U = 38,067.5, $p < 0.001$; Table 2). Malignancy rates rose monotonically with age, from 0% in the < 20-year group to 84.6% in the ≥ 70-year group ($\chi^2=163.7$, df = 6, $p < 0.001$). Large tumor size at presentation was more common among malignant cases (median 5.0 vs. 3.5 cm; $p < 0.001$), with T3/T4 tumors carrying the highest malignancy rate (55.1% vs. 16.4% for T1; $\chi^2=25.4$, df = 2, $p < 0.001$). Confirmed metastasis was exclusively associated with malignant outcomes ($\chi^2=110.2$, $p < 0.001$). Laterality was significantly associated with outcome ($\chi^2=20.5$, df = 4, $p = 0.0004$), driven largely by the very low malignancy rate in bilateral lesions (3.6%), which predominantly represent fibroadenoma. Sex was not significantly associated with outcome ($\chi^2=1.41$, $p = 0.235$).

Table 2
Bivariate Comparison of Clinical Characteristics by Outcome (N = 437)

Characteristic	Benign (n = 262)	Malignant (n = 175)	Test Statistic	p-value
Age (years): Median (IQR)	25 (20–36)	45 (36–57)	U = 38,067.5	< 0.001
Tumor Size (cm): Median (IQR)	3.5 (2.4–5.0)	5.0 (3.5–6.0)	U = 22,041.0	< 0.001
Sex: Female	253 (96.6%)	173 (98.9%)	χ^2=1.41	0.235
Age Group			χ^2=163.7, df = 6	< 0.001
<20 years	59 (100%)	0 (0%)		
20–29 years	103 (90.4%)	11 (9.6%)		
30–39 years	50 (54.3%)	42 (45.7%)		
40–49 years	23 (37.7%)	38 (62.3%)		
50–59 years	8 (16.0%)	42 (84.0%)		
60–69 years	13 (40.6%)	19 (59.4%)		
≥70 years	4 (15.4%)	22 (84.6%)		
Tumor Size Category			χ^2=25.4, df = 2	< 0.001
≤2 cm (T1)	51 (83.6%)	10 (16.4%)		
2–5 cm (T2)	118 (61.1%)	75 (38.9%)		
>5 cm (T3/T4)	53 (44.9%)	65 (55.1%)		
Confirmed Metastasis: Yes	0 (0%)	44 (100%)	χ^2=110.2	< 0.001
Laterality			χ^2=20.5, df = 4	0.0004
Left	99 (54.7%)	82 (45.3%)		
Right	104 (57.1%)	78 (42.9%)		
Bilateral	27 (96.4%)	1 (3.6%)		

Mann–Whitney U test for continuous variables; chi-squared test (Yates' correction) for categorical variables. $p < 0.05$ considered statistically significant.

3.6 Multivariable Logistic Regression and Model Performance

Complete-case restriction reduced the analytic sample from 437 (definitive outcomes) to 369 cases, including 149 malignant events (40.4%); 68 cases (15.6%) were excluded due to missing tumor size data. Age and tumor size were independently associated with malignant outcome (Table 3). Each

additional year of age was associated with an 11% increase in the adjusted odds of malignancy (OR 1.107, 95% CI: 1.083–1.131, p < 0.001). Compared to T2 tumors (reference), T3/T4 tumors (> 5 cm) were associated with significantly higher adjusted odds (OR 1.90, 95% CI: 1.06–3.39, p = 0.030), while T1 tumors (≤ 2 cm) were inversely associated with malignant outcome (OR 0.15, 95% CI: 0.06–0.41, p < 0.001). Despite relying on only two routinely available variables, the model demonstrated excellent discrimination (AUC = 0.876, 95% CI: 0.841–0.909; Nagelkerke R^2=0.493). The HL test suggested some degree of calibration mismatch (χ^2=22.6, df = 8, p = 0.004); given the exploratory objective and parsimonious predictor set, the model was retained primarily for characterizing associations rather than individual risk prediction.

Table 3
Multivariable Logistic Regression: Factors Independently Associated with Malignant Outcome (N = 369 Complete Cases)

Variable	Adjusted OR	95% CI Lower	95% CI Upper	p-value
Age (per year, continuous)	1.107	1.083	1.131	< 0.001
Tumor Size: ≤2 cm (T1) vs. 2–5 cm (T2)*	0.153	0.058	0.405	< 0.001
Tumor Size: >5 cm (T3/T4) vs. 2–5 cm (T2)*	1.898	1.064	3.385	0.030
Constant	0.016	0.007	0.037	< 0.001
Model Performance				
N (complete cases)†	369			
Malignant events	149 (40.4%)			
AUC (95% bootstrap CI)	0.876	0.841	0.909	—
Nagelkerke R^2	0.493			
Hosmer–Lemeshow χ^2 (df = 8)‡	22.6			0.004

*Reference: 2–5 cm (T2). OR=Odds Ratio; CI=Confidence Interval. Metastasis excluded due to quasi-complete separation. †Complete-case restriction reduced sample from 437 to 369; 68 cases excluded due to missing tumor size (15.6%). ‡HL p = 0.004 The model demonstrated excellent discrimination (AUC = 0.876), although calibration was imperfect according to the Hosmer–Lemeshow test. Therefore, the model should be interpreted primarily as identifying associations rather than providing individualized risk prediction.

3.7 Sensitivity Analyses

All four sensitivity analyses produced consistent findings (Fig. 2). Restricting to females (SA1; n = 361) yielded age OR 1.108 (95% CI: 1.084–1.132) and T3/T4 OR 2.12 (95% CI: 1.17–3.81; p = 0.013). Excluding extreme size outliers (SA2; n = 367) did not materially alter estimates. Restricting to adults ≥ 18 years (SA3; n = 340) confirmed minimal influence of pediatric cases. Using the broadened outcome

definition (SA4; n = 395, AUC = 0.848), the inverse association with T1 tumors was attenuated but remained significant (OR 0.29; 95% CI: 0.14–0.64; p = 0.002), and associations with age and T3/T4 remained stable. The uniformity of findings across all four scenarios reinforces the robustness of the primary results.

4. DISCUSSION

This retrospective study provides a five-year characterization of the histopathological spectrum of breast lesions from Mogadishu, Somalia. Among 522 consecutive cases, fibroadenoma was the dominant diagnosis (33.1%), consistent with reports from other sub-Saharan African populations in which fibroadenoma accounts for a substantial proportion of breast surgical specimens among young women [5,6]. The high proportion in the under-40 age groups reflects the young demographic structure of the Somali population and the common presentation of benign breast disease in the reproductive-age cohort.

The predominance of fibroadenoma in this cohort is also consistent with African histopathology series in which young women contribute a large share of breast biopsies and excisions. This finding should not be interpreted as reducing the importance of malignancy; rather, it emphasizes the dual workload of breast pathology services in low-resource settings: reassuring and managing a high volume of benign lesions while also detecting malignancies early enough for curative treatment [19–23,28–31].

The limited availability of ER, PR, HER2, and Ki-67 data is one of the most important findings from a health-system perspective. Multicountry African studies have shown that receptor-defined subtype distribution varies across countries and that treatment planning is compromised when IHC is unavailable or inconsistently reported [32–36,41–43]. For Somalia, investment in standardized tissue handling, synoptic reporting, receptor testing, and linkage between pathology, surgery, oncology, and radiology records would substantially improve both patient care and the quality of future research [15–18,37–40].

The study findings align with the World Health Organization Global Breast Cancer Initiative, which emphasizes health promotion for early detection, timely diagnosis, and completion of treatment as the three core pillars of breast cancer control [12]. In Mogadishu and similar low-resource urban settings, an achievable first step may be strengthening clinical breast assessment, rapid biopsy referral, standardized pathology reports, and targeted awareness for women aged 35 years and older, while building the infrastructure required for stage documentation and receptor-based treatment decisions [12,14,40–45].

The malignancy prevalence of 40.1% among evaluable cases must be interpreted within the context of a pathology-referred population. Invasive ductal carcinoma/NST was the predominant malignant subtype (~ 54% of malignant cases), consistent with global patterns in which IDC/NST constitutes approximately 70–75% of all invasive breast cancers [7]. The substantial proportion of cases with large tumor size at presentation (median 5.0 cm in malignant cases) is consistent with delayed care-seeking and limited early detection access — rather than confirmed advanced stage, since pathological staging was available for fewer than 13% of the cohort.

Age was the most strongly associated variable with malignant outcome in multivariable analysis. Despite relying on only two routinely available variables, the model demonstrated excellent discrimination (AUC = 0.876, 95% CI: 0.841–0.909). The median age at malignant diagnosis was 45 years, approximately a decade younger than the median in high-income countries (~ 62 years) [8], consistent with the earlier-onset pattern documented across sub-Saharan Africa [9]. These findings support heightened awareness and earlier clinical evaluation among women aged approximately 35 years and older, though a nationally prescribed screening threshold would require broader health system evidence beyond the scope of this study.

Tumor size was the second variable independently associated with malignancy, with T3/T4 tumors (> 5 cm) carrying nearly twice the adjusted odds and T1 tumors (≤ 2 cm) being inversely associated. This pattern was consistent across all four sensitivity analyses (Fig. 2), including when the outcome definition was broadened (SA4), where the inverse association with T1 tumors was attenuated to OR 0.29 but remained statistically significant, confirming the clinical informativeness of tumor size at presentation as a discriminating characteristic in this setting.

Regarding model calibration: the HL test suggested some degree of calibration mismatch (χ^2=22.6, p = 0.004), which is not uncommon in parsimonious models with skewed predictor distributions [10]. Given the study's primary aim of characterizing associations rather than generating individual risk scores, the model was appropriately retained for its discriminative value. Future studies with larger samples and additional clinical variables — including IHC markers and staging data — could develop and formally validate a calibrated risk prediction tool.

Several limitations should be acknowledged. The near-universal absence of IHC receptor data (ER, PR, HER2, Ki67 each available in < 2% of cases) reflects the current state of pathological infrastructure in the study setting and represents a significant evidence gap. Molecular subtype analysis, central to treatment planning in high-income settings, is not yet feasible from routine data in this context. Additional limitations include the retrospective design with reliance on free-text diagnoses requiring standardization, the complete-case analytic restriction excluding 68 cases (15.6%) primarily due to missing tumor size, and geographic restriction to Mogadishu — predominantly a single tertiary hospital — which may not represent the distribution of breast lesions in rural populations or other regions of Somalia.

5. CONCLUSION

This study provides important baseline data on the histopathological spectrum of breast lesions and factors associated with malignancy in Mogadishu, Somalia, drawn from five years of secondary pathological records. Fibroadenoma dominates among younger women, while malignancy — predominantly invasive ductal carcinoma — presents at an older age and with larger tumor size. Age and tumor size at presentation were independently associated with malignant outcome, yielding a model with excellent discriminative performance (AUC = 0.876, 95% CI: 0.841–0.909). These findings support

strengthening timely diagnostic evaluation and pathology services and earlier clinical evaluation among women aged approximately 35 years and older as well as investment in standardized histopathological reporting — including IHC capacity — in Somalia and comparable low-resource settings. Future prospective studies incorporating molecular profiling, complete staging, treatment, and survival data are needed to guide evidence-based, context-appropriate breast cancer management protocols.

Declarations

FUNDING

This research was funded by SIMAD university through institutional funding. The funding provided support for the conduct of the research. The funder had no role in the study design, data collection, data analysis, data interpretation, manuscript preparation, or the decision to submit the manuscript for publication

AUTHOR CONTRIBUTIONS

MSY (Mohamed Serar Yusuf): Conceptualization, study design, data curation, formal analysis, methodology, software, visualization, writing—original draft, and writing—review and editing.

MHA (Mohamed Hussein Adam): Conceptualization, supervision, validation, writing—review and editing, and critical revision of the manuscript for important intellectual content.

AAA (Abdkerem Abdulahi Abdi) : Supervision, validation, project administration, writing—review and editing, and critical revision of the manuscript for important intellectual content

RMA (Rahma Mohamed Ali): Data collection, data curation, and writing—review and editing.

All authors have read and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

Conflicts of Interest

The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Ethical Approval and Consent to Participate

Ethical approval was obtained from the Institutional Review Board of SIMAD University, approval number EC000245, dated 27 June 2026. Administrative permission was also obtained from the participating healthcare facilities. Because this was a retrospective study using de-identified histopathological records, the requirement for individual informed consent was waived by the ethics committee.

The study utilized retrospective, de-identified histopathological records; accordingly, the requirement for individual patient informed consent was waived by the respective ethics review bodies, consistent with the retrospective and non-interventional nature of the study. All data were anonymized prior to analysis and handled confidentially throughout the study in accordance with applicable data protection principles. Findings are reported only in aggregate form to ensure participant privacy and confidentiality.

Data Availability Statement

The data that support the findings of this study are not publicly available due to institutional data governance requirements and participant confidentiality considerations. De-identified aggregate data may be made available upon reasonable written request to the corresponding author, subject to approval by the institutional ethics committee and relevant data governance authorities of the participating facilities.

ACKNOWLEDGEMENTS

The authors used Paperpall AI (https://paperpal.com) for language editing to improve clarity and readability. This assistance was limited to linguistic refinement only; study design, data analysis, interpretation, and all conclusions remain the responsibility of the authors.

The authors also extend their heartfelt appreciation to the management and staff of Nova Diagnostic Center, Mogadishu, Somalia, which is a hospital-based, and private center, for their cooperation and assistance in facilitating access to pathological records and laboratory data essential for this study.

Furthermore, the authors acknowledge all participating hospitals and healthcare facilities in Mogadishu, Somalia, for their support, collaboration, and commitment to improving clinical research and patient care.

Finally, the authors are deeply grateful to the patients whose anonymized and de-identified clinical and pathological data contributed to this research. Their contribution has made this study possible and will help advance medical knowledge and healthcare practice in Somalia.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71(3):209-249.
2. Jedy-Agba E, McCormack V, Adebamowo C, dos Santos Silva I. Stage at diagnosis of breast cancer in sub-Saharan Africa: a systematic review and meta-analysis. Lancet Glob Health. 2016;4(12):e923-e935.
3. Anderson BO, Ilbawi AM, El Saghir NS. Breast cancer in low- and middle-income countries: a shifting tide in global health. Breast J. 2015;21(1):111-118.

4. Ali FM, Mohamed NH, Nur AA, et al. Challenges of cancer care in Somalia: a call to action. Ecancermedicalscience. 2020;14:1025.
5. Ihekwaba FN. Breast disease in Nigerian women. J R Coll Surg Edinb. 1994;39(1):7-10.
6. Gakwaya A, Kiguli-Malwadde E, Kavuma A, et al. Cancer of the breast: 5-year experience in a Uganda Cancer Institute. Afr Health Sci. 2008;8(1):43-47.
7. WHO Classification of Tumours Editorial Board. Breast Tumours. 5th ed. Lyon: International Agency for Research on Cancer; 2019.
8. Howlader N, Noone AM, Krapcho M, et al, editors. SEER Cancer Statistics Review, 1975-2018. Bethesda (MD): National Cancer Institute; 2021.
9. Adeloye D, Sowunmi OY, Jacobs W, et al. Estimating the incidence of breast cancer in Africa: a systematic review and meta-analysis. J Glob Health. 2018;8(1):010419.
10. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology. 2010;21(1):128-138.
11. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, Jemal A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74(3):229-263.
12. World Health Organization. Global Breast Cancer Initiative implementation framework: assessing, strengthening and scaling-up of services for the early detection and management of breast cancer. Geneva: World Health Organization; 2023.
13. International Agency for Research on Cancer. Global Cancer Observatory: Somalia fact sheet 2022. Lyon: IARC; 2024.
14. International Agency for Research on Cancer. Global Cancer Observatory: Eastern Africa fact sheet 2022. Lyon: IARC; 2024.
15. Tahtabasi M, Mohamud Abdullahi I, Kalayci M, Gedi Ibrahim I, Er S. Cancer incidence and distribution at a tertiary care hospital in Somalia from 2017 to 2020: an initial report of 1306 cases. Cancer Manag Res. 2020;12:8599-8611.
16. Walz L, James C, Mahamud R, et al. Knowledge, attitudes and practices concerning breast cancer, cervical cancer and screening among healthcare professionals and students in Mogadishu, Somalia: a cross-sectional study. Ecancermedicalscience. 2022;16:1455.
17. Osman AA, Ibrahim IG, Mohamed AH, et al. Patterns of histopathology presentation of breast cancer in a tertiary hospital in Somalia. Int Cancer Res J. 2025;1(1):1-8.
18. Nusretoglu R, et al. Breast lesions and cancer staging in Somalia: a retrospective pathology-based study from a regional referral centre. J Health Sci Med. 2026;9(1):1-8.
19. Olu-Eddo AN, Ugiagbe EE. Benign breast lesions in an African population: a 25-year histopathological review of 1864 cases. Niger Med J. 2011;52(4):211-216.
20. Okoth C, Galukande M, Jombwe J, Wamala D. Benign proliferative breast diseases among female patients at a sub-Saharan Africa tertiary hospital: a cross sectional study. BMC Surg. 2013;13:9.

21. Njeze GE. Breast lumps: a 21-year single-center clinical and histological analysis. Niger J Surg. 2014;20(1):38-41.
22. Anyikam A, Nzegwu MA, Ozumba BC, Okoye I, Olusina DB. Benign breast lesions in Eastern Nigeria. Saudi Med J. 2008;29(2):241-244.
23. Akinnibosun-Raji HO, Omiyi DO, Olowoyeye OA, et al. Correlation of sonographic findings and histopathological diagnosis of breast lesions in a tertiary health facility. J West Afr Coll Surg. 2022;12(4):80-86.
24. Ibrahim EH, Abdalla AA, Elhassan MM, et al. Histopathological profile of different breast lesions in a tertiary care hospital. Cureus. 2024;16(5):e60516.
25. Galukande M, Wabinga H, Mirembe F, Karamagi C, Asea A. Molecular breast cancer subtypes prevalence in an indigenous sub-Saharan African population. Pan Afr Med J. 2014;17:249.
26. Sayed S, Moloo Z, Wasike R, et al. Is breast cancer from sub-Saharan Africa truly receptor poor? Prevalence of ER/PR/HER2 in breast cancer from Kenya. Breast. 2014;23(5):591-596.
27. Huo D, Ikpatt F, Khramtsov A, et al. Population differences in breast cancer: survey in indigenous African women reveals over-representation of triple-negative breast cancer. J Clin Oncol. 2009;27(27):4515-4521.
28. Galukande M, Wabinga H, Mirembe F. Breast cancer survival experiences at a tertiary hospital in sub-Saharan Africa: a cohort study. World J Surg Oncol. 2015;13:220.
29. Kantelhardt EJ, Zerche P, Mathewos A, et al. Breast cancer survival in Ethiopia: a cohort study of 1,070 women. Int J Cancer. 2014;135(3):702-709.
30. Tenge CN, Kuremu RT, Buziba NG, Patel K, Were PA. Burden and pattern of cancer in Western Kenya. East Afr Med J. 2009;86(1):7-10.
31. Pace LE, Dusengimana JM, Hategekimana V, et al. Benign and malignant breast disease at Rwanda's first public cancer referral center. Oncologist. 2016;21(5):571-575.
32. Uyisenga JP, Butera Y, Debit A, Josse C, Nakimenyi M, Remacle J, et al. Prevalence of histological characteristics of breast cancer in Rwanda in relation to age and tumor stages. Horm Cancer. 2020;11(5-6):240-249.
33. Popli P, Abebe T, Bhatia A, et al. Receptor-defined breast cancer in five East African countries and its implications for treatment: systematic review and meta-analysis. JCO Glob Oncol. 2021;7:289-301.
34. Onyia AF, Tolu LB, Rulisa S, et al. Breast cancer phenotypes in Africa: a scoping review and meta-analysis. JCO Glob Oncol. 2023;9:e2300135.
35. McKenzie F, Zietsman A, Galukande M, et al. African Breast Cancer-Disparities in Outcomes (ABC-DO): protocol of a multicountry mobile health prospective study of breast cancer survival in sub-Saharan Africa. BMJ Open. 2016;6(8):e011390.
36. McCormack V, McKenzie F, Foerster M, et al. Breast cancer survival and survival gap apportionment in sub-Saharan Africa: the African Breast Cancer-Disparities in Outcomes cohort study. Lancet Glob Health. 2020;8(9):e1203-e1212.

37. Kim J, Macharia PM, McCormack V, Foerster M, Galukande M, Joffe M, et al. Geospatial disparities in survival of patients with breast cancer in sub-Saharan Africa from the African Breast Cancer-Disparities in Outcomes cohort: a prospective cohort study. Lancet Glob Health. 2024;12(7):e1111-e1119.
38. Unger-Saldana K. Challenges to the early diagnosis and treatment of breast cancer in developing countries. World J Clin Oncol. 2014;5(3):465-477.
39. Foerster M, McKenzie F, Zietsman A, et al. Dissecting the journey to breast cancer diagnosis in sub-Saharan Africa: findings from the multicountry ABC-DO cohort study. Int J Cancer. 2021;148(2):340-351.
40. Cazap E, Magrath I, Kingham TP, Elzawawy A. Structural barriers to diagnosis and treatment of cancer in low- and middle-income countries: the urgent need for scaling up. J Clin Oncol. 2016;34(1):14-19.
41. Vanderpuye V, Grover S, Hammad N, et al. An update on the management of breast cancer in Africa. Infect Agent Cancer. 2017;12:13.
42. Mutebi M, Anderson BO, Duggan C, et al. Breast cancer treatment: a phased approach to implementation. Cancer. 2020;126 Suppl 10:2365-2378.
43. Kingham TP, Alatise OI, Vanderpuye V, et al. Treatment of cancer in sub-Saharan Africa. Lancet Oncol. 2013;14(4):e158-e167.
44. El Saghir NS, Adebamowo CA, Anderson BO, et al. Breast cancer management in low-resource countries: consensus statement from the Breast Health Global Initiative. Breast. 2011;20 Suppl 2:S3-S11.
45. World Health Organization. Guide to cancer early diagnosis. Geneva: World Health Organization; 2017.

Figures

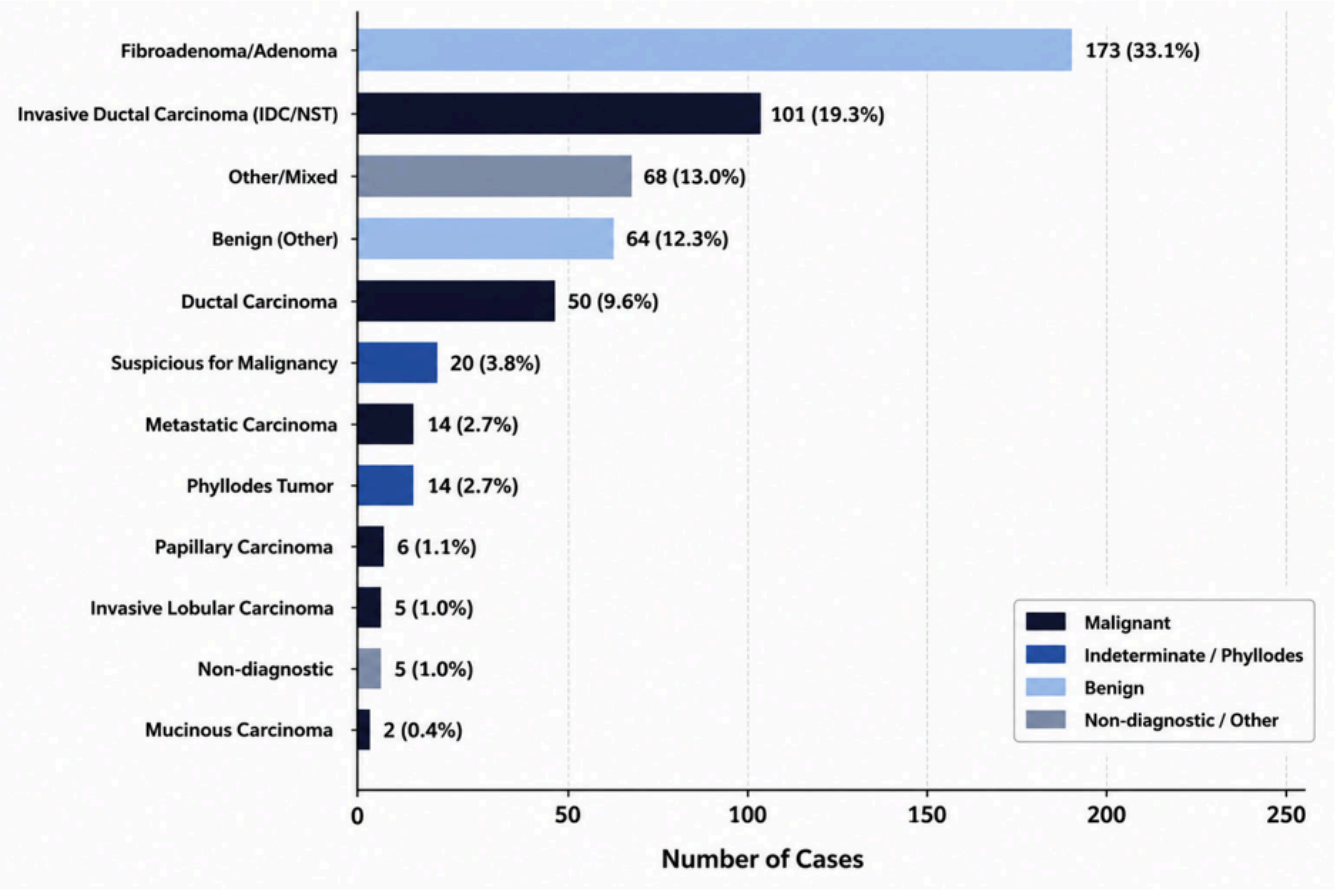


Figure 1

Histopathological Spectrum of Breast Diagnoses (N=522, Mogadishu, Somalia, July 2021–December 2025). Bars represent number of cases per histopathological category, ordered by frequency. Color denotes diagnostic grouping: dark navy = malignant; mid-blue = indeterminate/phyllodes; light blue = benign; grey = non-diagnostic/other. Counts and proportions of total (N=522) are annotated.

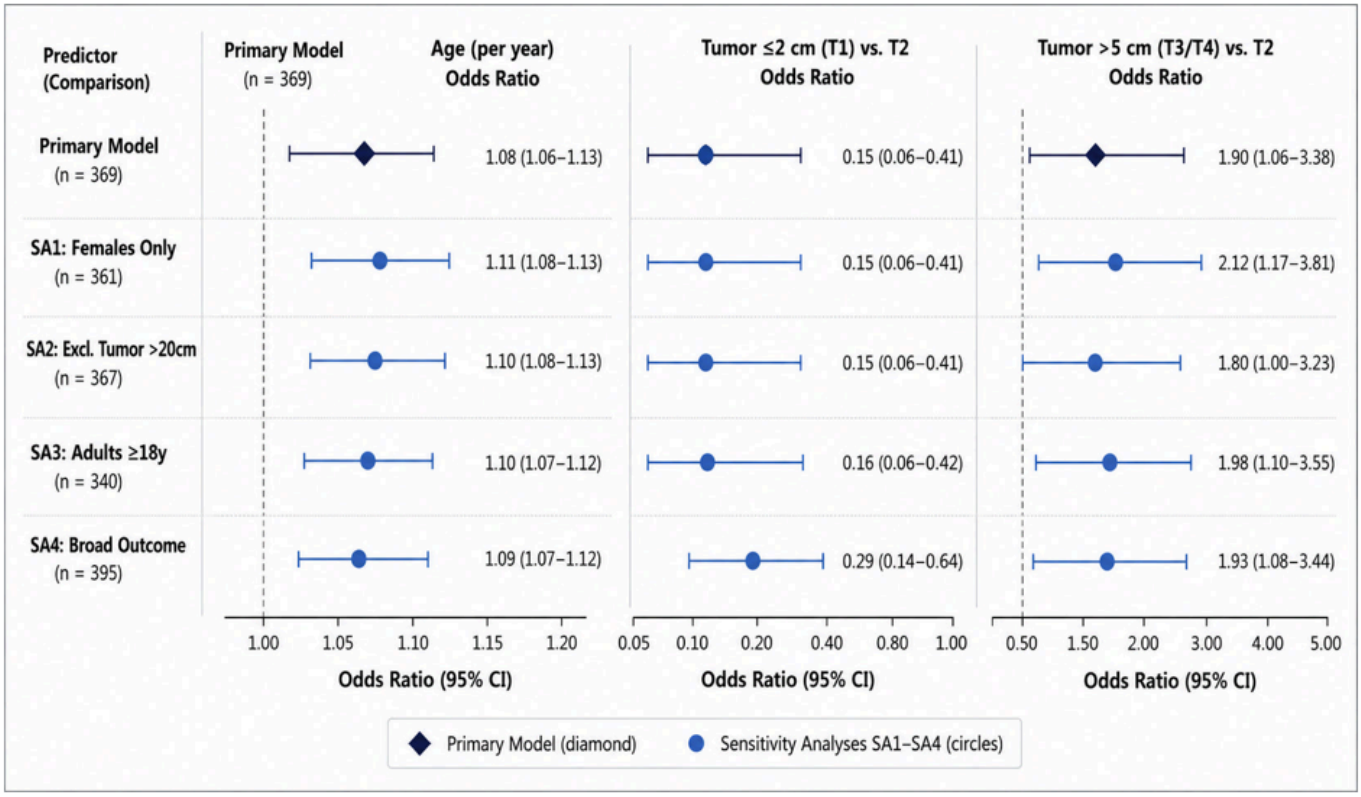


Figure 2

Forest Plot of Sensitivity Analyses: Consistency of Associations Across Alternative Analytical Scenarios. Each panel shows odds ratios (OR) with 95% confidence intervals. Primary model shown as a diamond (dark navy); sensitivity analyses SA1–SA4 as circles (mid-blue). Left: Age per year. Centre: Tumor ≤2 cm (T1) vs. T2. Right: Tumor >5 cm (T3/T4) vs. T2. Dashed vertical line = OR 1.0. SA1=females only; SA2=exclude tumor >20 cm; SA3=adults ≥18y; SA4=broadened outcome (Suspicious/Atypical coded as malignant).