



Breast Cancer Screening Utilization Patterns and Detection in Brazil's Private Health Sector: Evidence Using Artificial Intelligence, 2014–2023

2805

Andrade M. V., Noronha K., Ribeiro L. C., Kelles S. B., Cherchiglia M., Julião N. A., Carvalho L., Bersan S. A. L., Rigotti M. L., Colares F., Couto F., Silva N. G. P., Benner P., Sanna C., Amorin P. H., Carvalho C., Bracarense H., Borin M., Michel M.



Background

In Brazil, malignant neoplasms are a leading cause of death, and early detection is crucial. However, screening protocols are inconsistently followed, and the widespread availability of services often results in unnecessary overuse, particularly in private sector.

Objective: To identify and characterize mammography screening utilization profiles using longitudinal administrative healthcare data from Brazil's private health sector, and to assess their alignment with screening guidelines.

Data

58,332 women | 10-year balanced longitudinal panel

We analyze administrative healthcare data from a large private health insurer in Brazil, covering the period from 2014 to 2023. The study population consists of a balanced panel of 58,332 women aged 18 to 69 years who remained continuously enrolled throughout the study period. The dataset includes demographic characteristics, health plan information, and detailed records of healthcare utilization. Breast cancer diagnosis is identified using ICD-10 codes and validated against an independent oncology database, ensuring high accuracy.

Methods

The empirical strategy proceeds in three steps. First, a neural network model predicts breast cancer diagnosis based on utilization patterns of screening, diagnostic, and treatment procedures, achieving high predictive performance with accuracy above 95%.

Second, individual screening trajectories are represented as time series and transformed into the frequency domain using the Discrete Fourier Transform, capturing periodic patterns of utilization. These features are projected into a lower-dimensional space using t-SNE and clustered using DBSCAN to identify groups with similar temporal behavior.

Finally, clusters are aggregated into seven macro-groups using latent profile analysis, based on utilization intensity, frequency, age distribution, and cancer prevalence.

Results

A total of 53 mammography utilization clusters were identified over the ten-year period. Cluster separation captured not only differences in overall use, but also distinct temporal patterns of screening behavior (Figure 1).

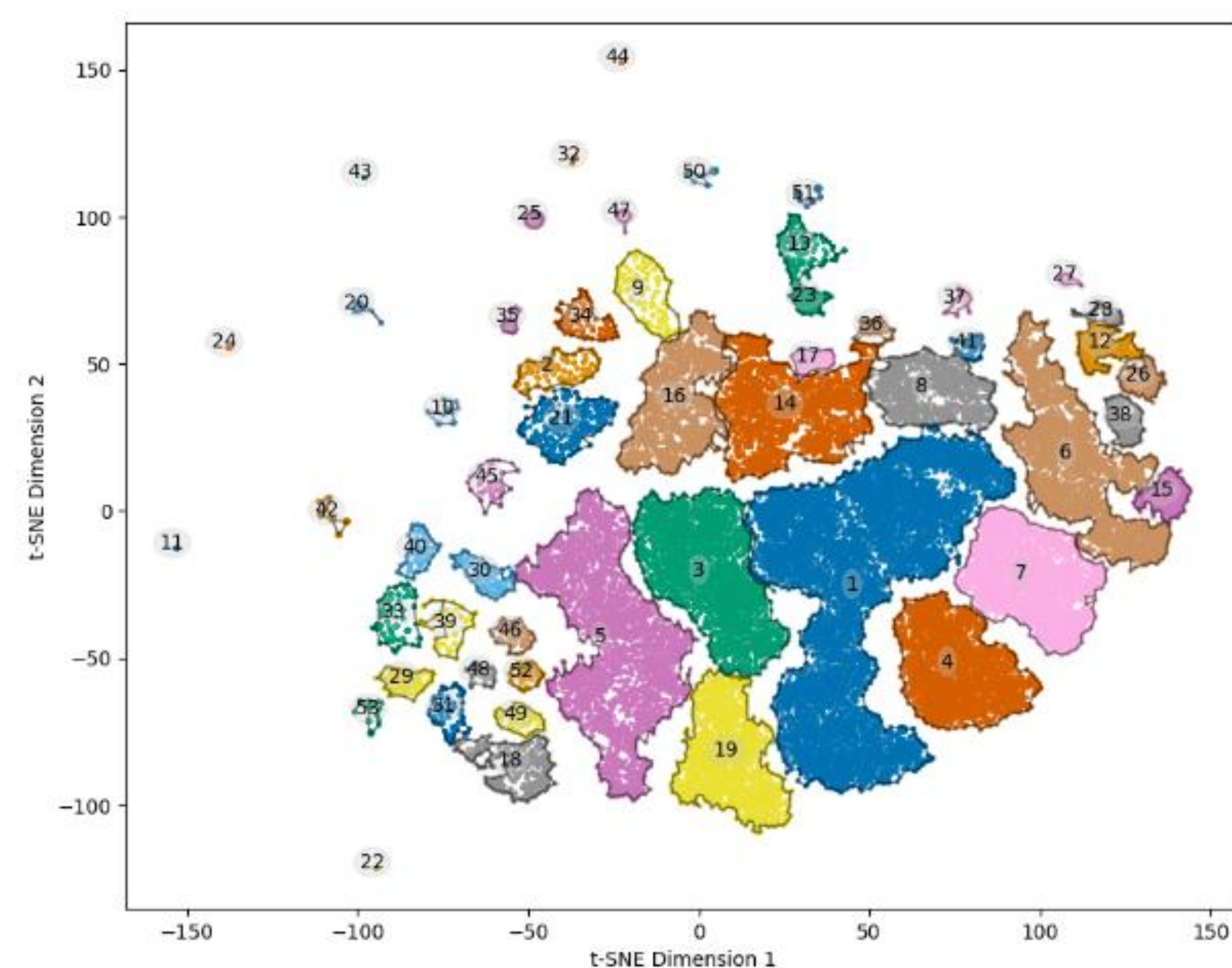


Figure 1. Fifty-three clusters of screening utilization patterns identified via DBSCAN on t-SNE embeddings.

The macro-groups reveal substantial heterogeneity in screening behavior (Figure 2). Approximately 43% of the population is classified as underuse, including both appropriate non-participation among younger women and insufficient screening among near-eligible individuals.

Around 31% exhibit guideline-consistent behavior, while 26% show above-guideline use, including frequent or repeated examinations often associated with higher cancer prevalence or follow-up care.

Macro-groups	Mean mammograms	Prop. of women with mammography	Mean age	Prevalence of cancer diagnosis
7	1.04	0.90	48.18	32.52%
6	0.95	0.91	48.83	15.38%
5	0.14	0.14	39.00	1.46%
4	0.36	0.36	43.96	3.40%
3	0.60	0.58	46.49	5.86%
2	0.10	0.10	36.81	1.84%
1	0.77	0.74	47.74	10.37%

Figure 2. Seven mammography screening macro-groups defined by utilization intensity, frequency, age, and cancer prevalence. Shading indicates relative levels across indicators. MG1 (12%): high participation above guidelines among mostly eligible women; MG2 (16%): minimal participation among predominantly non-eligible women; MG3 (31%): guideline-consistent screening with transitional age composition; MG4 (25%): sub-biennial participation among near-eligible women; MG5 (6%): low utilization driven by younger age composition; MG6 (8%): very high participation with repeated use and elevated risk; MG7 (1%): high utilization strongly driven by disease and follow-up care.

Discussion

This study shows that breast cancer screening behavior cannot be adequately described by aggregate utilization measures alone. By combining machine learning and time-series methods, we identify meaningful heterogeneity in screening trajectories, distinguishing appropriate non-use, potential underuse, guideline-consistent behavior, and high-intensity use.

These findings provide actionable insights for targeting screening interventions and improving the efficiency of breast cancer screening programs.

