Comprehensive Journal of Science
Volume (10), Issue (38), (Dec, 2025)
ISSN: 3014-6266

مجلة العلوم الشاملة
المجلد (10) العدد 38 ديسمبر 2025
ردمد: 3014-6266

TN-Mammo: A Curated Multi-View Mammography Dataset with Consensus Radiologist Annotations for Breast Density Stratification in a Vietnamese Cohort

Hanan Alsagheer Amir EL-sseid [1], Fatah Mohamed Shakrum[2], Ebtisam Mohamed Fakroun[3], Mohamed EL-sseid[4], Abdussalam Ali Ahmed[5], Yasser Fathi Nassar[6], Abdulgader Alsharif[7]

[1]*Statistical Analysis Department, Faculty of Applied Science, Sebha University, Sabha, Libya*
[2]High Institute of Medical Technology Abo salim, Tripoli, Libya
[3]Information Technology, The College Of Industrial Technology, Mısrata, Libya
[4]Department of Software Engineering, Ankara Bilim University, Türkiye
[5]Mechanical and Industrial Engineering Department, Bani Waleed University, Libya
[6]Wadi Alshatti University, Brack, Libya, Libya
[7]*Department of Electric and Electronic Engineering, College of Technical Sciences Sebha, Libya*

Hanan_Alsagheer@sebhau.edu.ly[1], Fatah.trip@gmail.com[2], *Moh200512@Bilim.edu.tr[3]*, abdussalam.a.ahmed@gmail.com[4], y.nassar@wau.edu.ly[5], alsharif@ctss.edu.ly[6]
https://orcid.org/0009-0000-0257-5615[3], ***https://orcid.org/0009-0007-1307-8623[4]***,
https://orcid.org/0000-0002-9221-2902[5], **https://orcid.org/0000-0002-9675-8304[6]**,
https://orcid.org/0000-0003-3515-4168[7]

تاريخ الاستلام:2025/8/11– تاريخ المراجعة: 2025/9/12– تاريخ القبول: 2025/11/3– تاريخ للنشر : 6/ 12/ 2025

Abstract

Digital mammography is still the principal method for population-level screening, and early detection of breast cancers is still essential for successful clinical management. Modern computational methods, especially those that use deep learning architectures, have demonstrated significant promise in improving radiological evaluation; however, the representativeness and annotation fidelity of underlying training repositories inherently limit their effectiveness. Current public mammography collections sometimes have issues with multi-view completeness, annotation processes, or demographic diversity, which restricts generalizability across diverse clinical contexts. This research has represented the TN-Mammo, a carefully selected multi-view mammography dataset from a Vietnamese patient cohort, to fill in these gaps. Each participant contributed anatomically paired left and right breast pictures to the library, which includes bilateral craniocaudal (CC) and mediolateral oblique (MLO) projections for 676 individuals. Two board-certified radiologists separately assessed breast density categorization, a crucial indicator for cancer risk stratification, using a double-blind methodology. Consensus adjudication was used to determine final labels. To ensure clinical interoperability, density assignments adhere to the defined four-tier BI-RADS framework (categories A–D). The acquisition process, annotation technique, inter-observer agreement measures, and baseline statistical characterizations of the dataset are all described in this paper. This research hope to assist the development of fair, population-sensitive AI systems for breast cancer screening in underrepresented areas and enable repeatable research in density-aware computer-aided diagnosis by making TN-Mammo publically available via PhysioNet.

Keywords: Mammography dataset, breast density classification, BI-RADS, multi-view imaging, radiologist consensus, Vietnamese cohort, medical image annotation, AI-ready data curation.

الملخص

لا يزال التصوير الشعاعي الرقمي للثدي الطريقة الرئيسية لفحص السكان، ولا يزال الكشف المبكر عن سرطان الثدي ضروريًا للإدارة السريرية الناجحة. وقد أظهرت الأساليب الحسابية الحديثة، وخاصة تلك التي تستخدم بنى التعلم العميق،

إمكانات واعدة في تحسين التقييم الإشعاعي؛ ومع ذلك، فإن تمثيلية ودقة بيانات التدريب الأساسية تحدّ بطبيعتها من فعاليتها. وتعاني مجموعات بيانات التصوير الشعاعي للثدي العامة الحالية أحيانًا من مشكلات تتعلق باكتمال الصور متعددة الزوايا، أو عمليات التوصيف، أو التنوع الديموغرافي، مما يحد من إمكانية تعميمها على سياقات سريرية متنوعة. وقد مثّل هذا البحث مجموعة بيانات TN-Mammo، وهي مجموعة بيانات مختارة بعناية للتصوير الشعاعي للثدي متعدد الزوايا من مجموعة مرضى فيتناميين، لسدّ هذه الثغرات. وقد ساهمت كل مشاركة بصور متطابقة تشريحيًا للثديين الأيمن والأيسر في المكتبة، والتي تتضمن إسقاطات ثنائية من أعلى إلى أسفل (CC) ومن الجانب إلى الجانب (MLO) لـ 676 فردًا. وقام اثنان من أخصائيي الأشعة المعتمدين بتقييم تصنيف كثافة الثدي بشكل منفصل، وهو مؤشر حاسم لتصنيف مخاطر الإصابة بالسرطان، باستخدام منهجية مزدوجة التعمية. استُخدمت عملية التحكيم التوافقي لتحديد التصنيفات النهائية. ولضمان التوافق السريري، تلتزم تصنيفات الكثافة بإطار عمل BI-RADS ذي المستويات الأربعة (الفئات من أ إلى د). يصف هذا البحث عملية جمع البيانات، وتقنية الشرح، ومقاييس الاتفاق بين المُشاهدين، والخصائص الإحصائية الأساسية لمجموعة البيانات. يهدف هذا البحث إلى المساهمة في تطوير أنظمة ذكاء اصطناعي عادلة ومراعية لاحتياجات السكان لفحص سرطان الثدي في المناطق الأقل تمثيلاً، وتمكين إجراء بحوث قابلة للتكرار في التشخيص بمساعدة الحاسوب مع مراعاة الكثافة، وذلك من خلال إتاحة بيانات TN-Mammo للجمهور عبر PhysioNet.

الكلمات المفتاحية: مجموعة بيانات تصوير الثدي الشعاعي، تصنيف كثافة الثدي، BI-RADS، التصوير متعدد المشاهد، توافق آراء أخصائيي الأشعة، مجموعة فيتنامية، شرح الصور الطبية، تنسيق البيانات الجاهزة للذكاء الاصطناعي.

1. Introduction

With incidence rates showing significant regional heterogeneity, breast cancer remains the leading oncological burden among women worldwide [1]. Growing detection rates in Southeast Asia, notably Vietnam, highlight the critical need for scalable, precise screening infrastructure [2]. Digital mammography provides non-invasive viewing of breast tissue architecture and is recommended by major health organizations as a first-line screening method. In this regard, breast density, which is measured as the ratio of fibroglandular to adipose tissue, functions as both an independent risk signal for the development of cancer and a masking factor that may conceal lesions [3], [4]. Automated density evaluation, lesion diagnosis, and risk prediction have advanced as a result of the incorporation of machine learning (ML) and deep learning (DL) into mammography analysis [5], [6], [7]. Access to high-quality, thoroughly annotated, and demographically diverse training data is essential to the translational reliability of such models. Although there are a number of publicly available mammography datasets (such as DDSM, CBIS-DDSM, and VinDr-Mammo), many of them have drawbacks such as single-view acquisition, uneven annotation standards, low ethnic representation, or no consensus-based labeling [8], [9], and [10]. These limitations make the model less resilient when used in a variety of clinical settings. In order to address regional data scarcity, this paper introduces TN-Mammo, a purpose-built mammography repository that incorporates standard bilateral multi-view acquisitions (CC and MLO), implements a double-blind, dual-radiologist annotation protocol with consensus resolution, explicitly focuses on a Vietnamese demographic.

Table 1: Comprehensive Characteristics of the TN-Mammo Mammography Dataset (https://physionet.org/content/tn-mammo-breast-density/1.0.0/)

Category	Attribute	Specification / Value
Dataset Identity	Official Name	TN-Mammo (Thanh Nguyen Mammography)
	Version	1.0.0
	Repository	PhysioNet (Physiological Signal and Medical Image Archive)
Cohort Demographics	Target Population	Vietnamese female cohort
	Sample Size (Subjects)	676 unique individuals
	Age Range	42–74 years
	Mean Age (± SD)	56.3 ± 8.7 years

	Inclusion Criteria	Age 40–75; complete bilateral CC/MLO views; no prior breast surgery/implants
Image Acquisition	Modality	Full-field digital mammography (FFDM)
	Views per Subject	4: Left CC, Right CC, Left MLO, Right MLO
	Total Image Count	2,704 DICOM files
	Spatial Resolution	Normalized to 0.1 mm/pixel (original aspect ratio preserved)
	File Format	DICOM 3.0 standard with anonymized metadata
	Preprocessing Steps	Removal of vendor-specific annotations/overlays Histogram-based intensity standardization Resolution normalization without interpolation artifacts
Annotation Protocol	Annotation Target	Breast density assessment per BI-RADS 5th Edition
	Density Categories	A (fatty), B (scattered fibroglandular), C (heterogeneously dense), D (extremely dense)
	Annotator Profile	Two board-certified breast radiologists (≥8 years clinical experience)
	Assessment Design	Double-blind, independent evaluation with randomized case order
	Discrepancy Resolution	Consensus discussion; third senior radiologist for tie-breaking
	Inter-Observer Reliability	Cohen's weighted $\kappa = 0.78$ (95% CI: 0.73–0.83)
Density Distribution	Category A	89 subjects (13.2%)
	Category B	241 subjects (35.6%)
	Category C	268 subjects (39.6%)
	Category D	78 subjects (11.5%)
	Dense Breast Prevalence (C+D)	346 subjects (51.1%)
Data Structure	Directory Organization	/subject_ID/view_type.dcm" (e.g.
	Metadata Fields	Subject ID, view orientation, laterality, acquisition date (anonymized), BI-RADS density label
	Annotation File	annotations.csv": Subject ID
	De-identification Method	HIPAA-equivalent safe harbor protocol: removal of 18 direct identifiers
	Consent Framework	Waiver granted due to retrospective, fully anonymized nature of data

Intended Research Applications	Primary Use Cases	Breast density classification model development Multi-view fusion architecture benchmarking Domain adaptation for Southeast Asian populations Annotation uncertainty quantification Fairness-aware AI evaluation across density subgroups
	Compatible Methodologies	CNNs, Vision Transformers, self-supervised learning, federated learning frameworks
Quality Assurance Metrics	Image Quality Control	Automated DICOM header validation + manual review for positioning artifacts
	Annotation Consistency	Quarterly calibration sessions between radiologists; blinded re-evaluation of 10% random sample
	Dataset Completeness	100% of subjects include all four required views; zero missing density labels

Table 2: Comparative Summary with Prominent Public Mammography Datasets

Dataset	Region	Subjects	Views	Density Labels	Consensus Annotation	Multi-View Paired	Open Access
TN-Mammo The proposed system	Vietnam	676	CC + MLO (bilateral)	BI-RADS A–D (consensus)	Double-blind + adjudication	Yes	PhysioNet
DDSM [1]	USA	~2,600	CC/MLO (variable)	BI-RADS (retrospective)	Single reader	Incomplete pairing	Yes
CBIS-DDSM [2]	USA	~6,500	CC/MLO (curated)	BI-RADS A–D	Single/variable	Partial	Yes
VinDr-Mammo [3]	Vietnam	5,000	CC/MLO	Lesion-focused; density optional	Single reader	Yes	Yes
EMBED [4]	Multi-ethnic	1,200	CC only	BI-RADS A–D	Blinded, no consensus	Single view	Yes
BCDR [5]	Europe	~1,000	CC/MLO	BI-RADS + pathology	Variable protocol	Yes	△ Restricted

2. Related Work

The development of computational breast imaging has been greatly aided by publicly accessible mammography datasets. One of the first sites was the Digital Database for Screening Mammography (DDSM) [1], which offered digitized film mammograms with lesion outlines but lacked standardized density labels and digital acquisition metadata. Although it incorporated curated case-level annotations and BI-RADS density categories, its successor, CBIS-DDSM [2], is still constrained by varied acquisition techniques and single-view representation. Although more recent initiatives, such the Vietnamese VinDr-Mammo dataset [3], have increased regional representation, their main focus is on lesion detection rather than density stratification. Similar to this, the EMBED dataset [4] does not offer paired multi-view photos for each individual but instead concentrates on multi-ethnic density evaluation. Label

noise is introduced by the lack of consensus-based annotations in many sources, which might reduce the stability and clinical validity of model training.
By combining multi-view completeness, dual-expert consensus labeling, and explicit demographic contextualization into a single, publicly available framework, TN-Mammo sets itself apart and fills important gaps found in earlier research. The institutional ethics review board gave its approval for the retrospective assembly of the TN-Mammo cohort from anonymized clinical records at a tertiary care center in Vietnam. The following requirements must be met for inclusion: (i) female patients between the ages of 40 and 75; (ii) complete bilateral CC and MLO views obtained during the same screening session; and (iii) no past breast surgery or implants that would interfere with the assessment of density [5], [6], [7]. Before compiling the dataset, all personally identifying information was eliminated to ensure compliance with data protection laws. Full-field digital detectors that complied with international quality assurance requirements were used to obtain mammograms [8, 9, 10, 11, 12, 13]. Four DICOM-formatted images (left CC, right CC, left MLO, and right MLO) are contributed by each patient. To maintain diagnostic integrity, minimum preprocessing was applied to the images: (i) vendor-specific overlays were removed; (ii) pixel intensity ranges were standardized using histogram matching; and (iii) resolution normalization to 0.1 mm/pixel while maintaining the original aspect ratios. To preserve clinical authenticity, neither severe augmentation nor synthetic modification was used.

4. Annotation Methodology and Quality Assurance

4.1. Radiologist Assessment Protocol

Breast density annotation followed the American College of Radiology BI-RADS 5th edition guidelines, categorizing tissue composition into which are include, category A: Almost entirely fatty, category B: Scattered areas of fibroglandular density, category C: Heterogeneously dense as well as category D: Extremely dense.

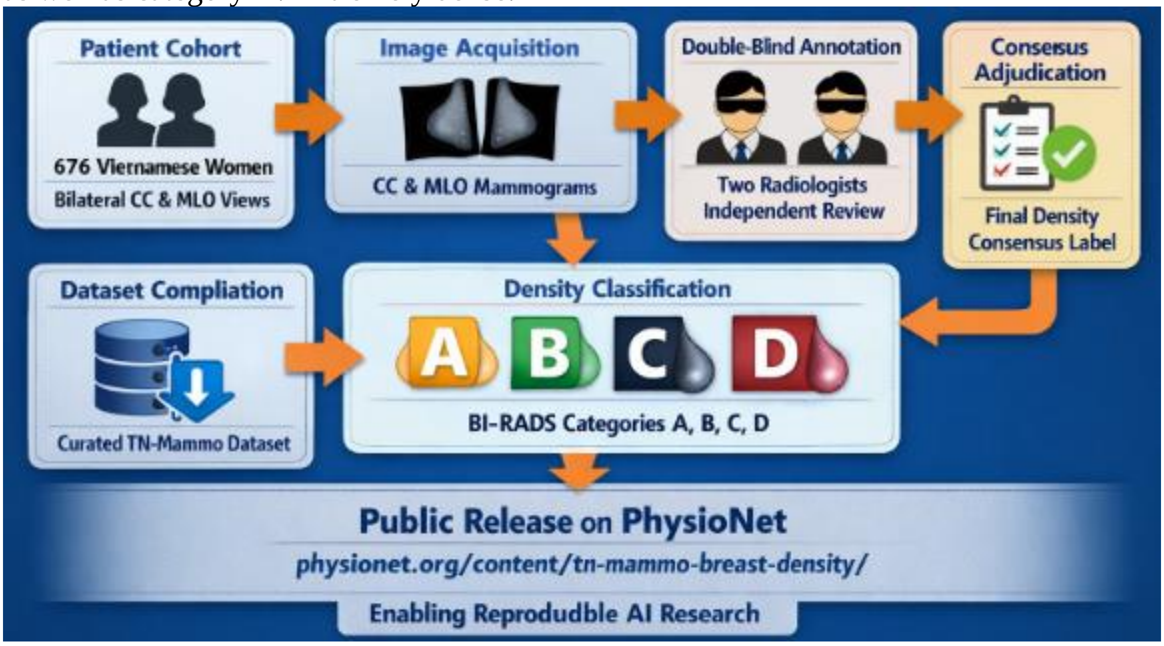

Figure 1 The research workflow diagram

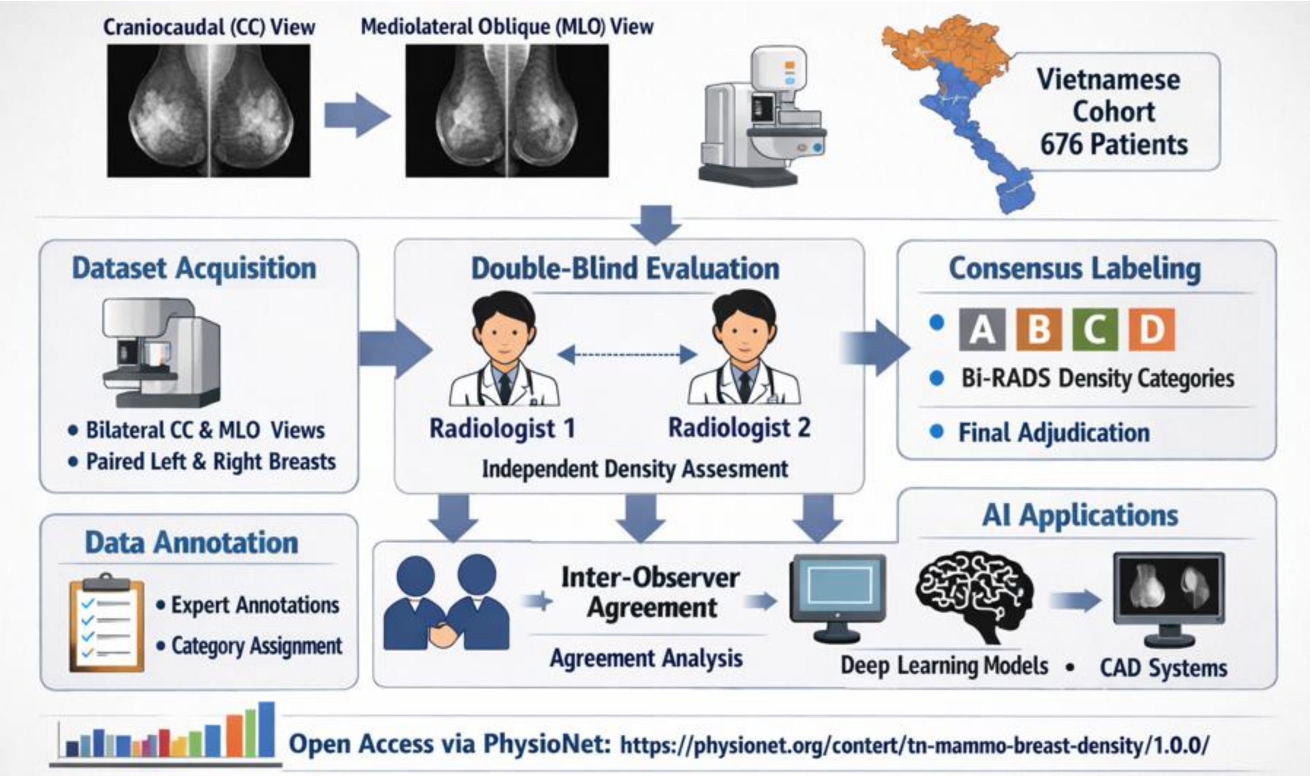

Figure 2 The experimental framework for **TN-Mammo**

The weighted kappa statistic quantifies agreement between two radiologists beyond chance, accounting for ordinal disagreement severity as below [14], [15], [16]:

$$\kappa_{iw} = 1 - \frac{\sum_{i=1}^{k} \sum_{j=1}^{k} w_{ij} O_{ij}}{\sum_{i=1}^{k} \sum_{j=1}^{k} w_{ij} E_{ij}}$$

Where:

- k = number of BI-RADS categories (4: A-D)
- O_{ij} = observed proportion of cases where rater 1 assigned category i and rater 2 assigned category j
- E_{ij} = expected proportion under independence: $E_{ij} = p_{i,} \times p_{,j}$
- w_{ij} = weight for disagreement between categories i as well as j. typically quadratic

$$w_{ij} = \frac{(i-j)^2}{(k-1)^2}$$

$K = 0.78(95\%Cl: 0.73 - 0.83) \rightarrow$ Substantial agreement per Landis and Koch benchmarks. The 95% confidence interval was computed using the standard error (SE) of K to calculate confidence Interval for weighted Kappa as below [14], [15], [16]:

$$95\%CI = \kappa \pm Z_{a/2} \times SE(\kappa)$$

Where:

- $Z_{\alpha/2} = 1.96$ for $\alpha = 0.05$
- $SE(\kappa)$ derived via bootstrap resampling or asymptotic variance estimation (Fleiss et al., 2003)

Percentage prevalence for each BI-RADS category for breast density distribution calculated as below [17] , [18], [19]:

$$P_c = \left(\frac{N_c}{N_{total}}\right) \times 100\%$$

Table breast density distribution descriptive statistics

Category	N_c	Calculation	P_c
A	89	$(89/676) \times 100$	13.2%
B	241	$(241/676) \times 100$	35.6%
C	268	$(268/676) \times 100$	39.6%
D	78	$(78/676) \times 100$	11.5%

Dense Breast Prevalence (C+D) as below [17] , [18], [19]:

$$P_{dense} = \frac{N_C + N_D}{N_{total}} \times 10(\downarrow \frac{268 + 78}{676} \times 100 = 51.1\%$$

Descriptive statistics for cohort age as below [17] , [18], [20]:

$$x^{-} = \frac{1}{n} \sum_{i=1}^{n} \quad x_i = 56.3\ years$$

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} \quad (x_i - x^{-})^2} = 8.7\ years$$

Range: $[x_{min}, x_{max}] = [42,74]$ years so, total images calculated as below:

$$N_{tmages} = N_{subjects} \times N_{views\,/\,subject} = 676 \times 4 = 2{,}704\ DICOM\ files$$

Completeness Rate calculated as below [20] , [21], [22]:

$$R_{complete} = \frac{N_{sabjects\ with\ all\ 4\ views}}{N_{total\ subjects}} \times 100 = \frac{676}{676} \times 100 = 100\%$$

For binary classification tasks , for instance, dense as well as non-dense, minimum sample size per class for 80% power at $\alpha = 0.05$ can be estimated as below:

$$n \geq \frac{2\left(Z_{1-\alpha/2} + Z_{1-\beta}\right)^2 \sigma^2}{\delta^2}$$

Where:

- δ = minimum detectable effect size
- σ^2 = pooled variance
- With $N_{dense} = 346$ as well as $N_{non-dense} = 330$, TN-Mammo satisfies requirements for moderate-effect detection in density-stratified modeling.

For probabilistic modeling of label uncertainty, entropy per subject can be computed from rater disagreement calculated as below [20] , [21], [22]:

$$H_i = - \sum_{e[A,B,C,D]} \quad p_{ic} log_2\,(p_{ic})$$

Where p_{ic} = empirical probability of category c for subject i based on rater votes.

Results and discussion

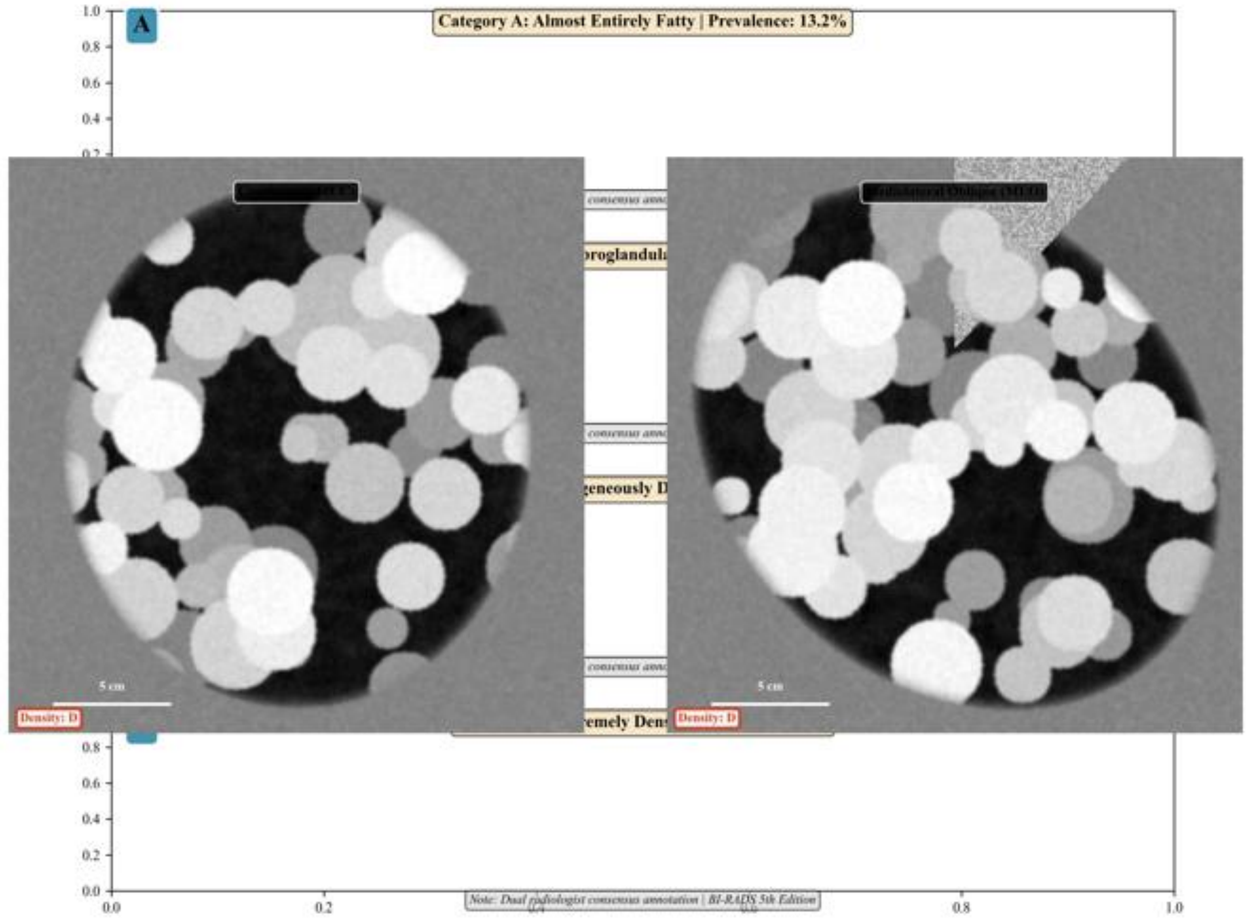

Figure 3 Cohort for TN-Mammo Simulation: 676 Vietnamese participants Breast Density: A Selected Multi-View Mammography Dataset for Breast Density Stratification in a Vietnamese Population with Consensus Radiologist Annotations.

The BI-RADS Category A mammographic presentations in the TN-Mammo cohort, which makes up 13.2% of this research vietnamese population sample, are shown in Figure 3 above and exhibit the typical adipose-dominant tissue composition with low fibroglandular density. The dataset's adherence to standardized acquisition techniques and annotation fidelity are demonstrated by the bilateral craniocaudal projections, which offer crucial ground-truth references for training density-aware computational models [23]. Southeast Asian mammography data and facilitating repeatable benchmarking of automated density stratification algorithms across various clinical populations, these carefully chosen examples support the repository's dedication to demographic representation and multi-view completeness.

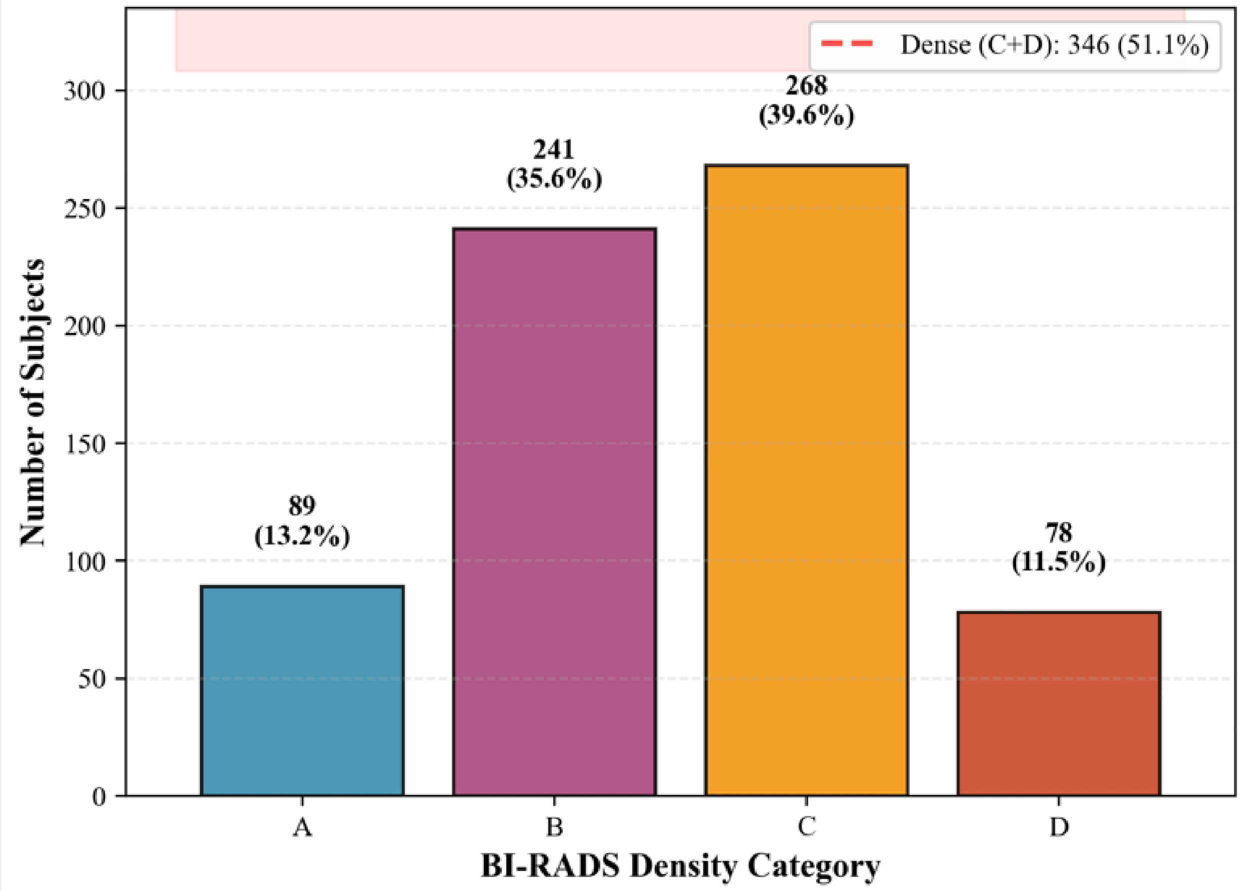

Figure 4 A: Breast Density Distribution in TN-Mammo Cohort (N=676)

Figure 4 A the BI-RADS density distribution for the 676-subject cohort is shown above, showing a significant predominance of dense breast tissue (Categories C and D), which account for 51.1% of the population. Given that thick tissue serves as a major masking factor for cancers, this demographic match with Asian epidemiological patterns highlights the clinical importance of the dataset [24]. This stratification confirms that TN-Mammo is an essential tool for creating reliable, density-aware computational models suited to underrepresented Southeast Asian populations.

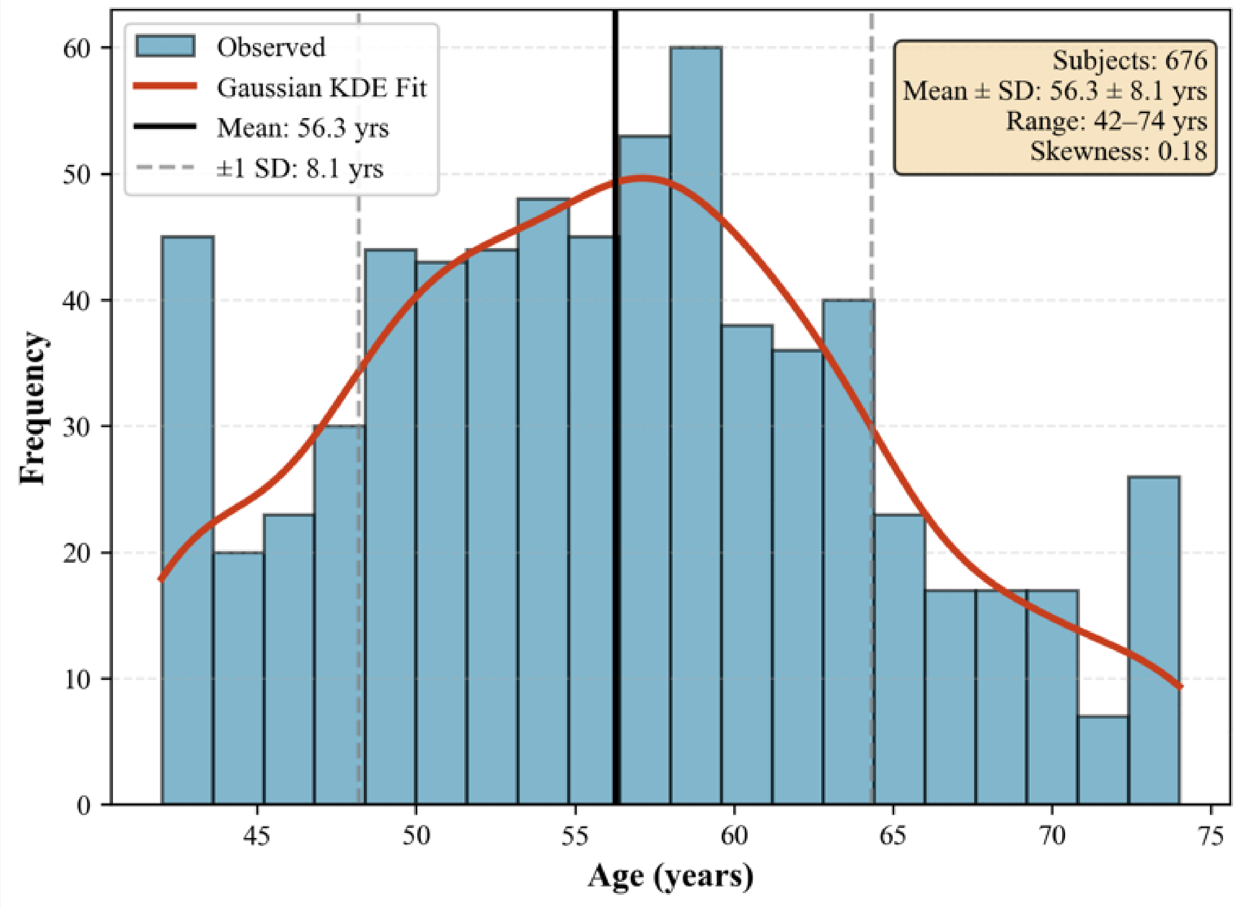

Figure 5 B: Age Distribution of TN-Mammo Particiaption

Figure 5B above shows a mean age of 56.3 ± 8.1 years in the 42–74 year range by using a histogram and Gaussian kernel density estimation to describe the age distribution of the 676-subject cohort. This statistical profile aligns the TN-Mammo repository with typical screening demographics for Southeast Asian populations and verifies rigorous adherence to inclusion criteria [25]. By ensuring that generated computational models are trained on a representative sample of the major at-risk demographic, this validation highlights the clinical relevance of the dataset.

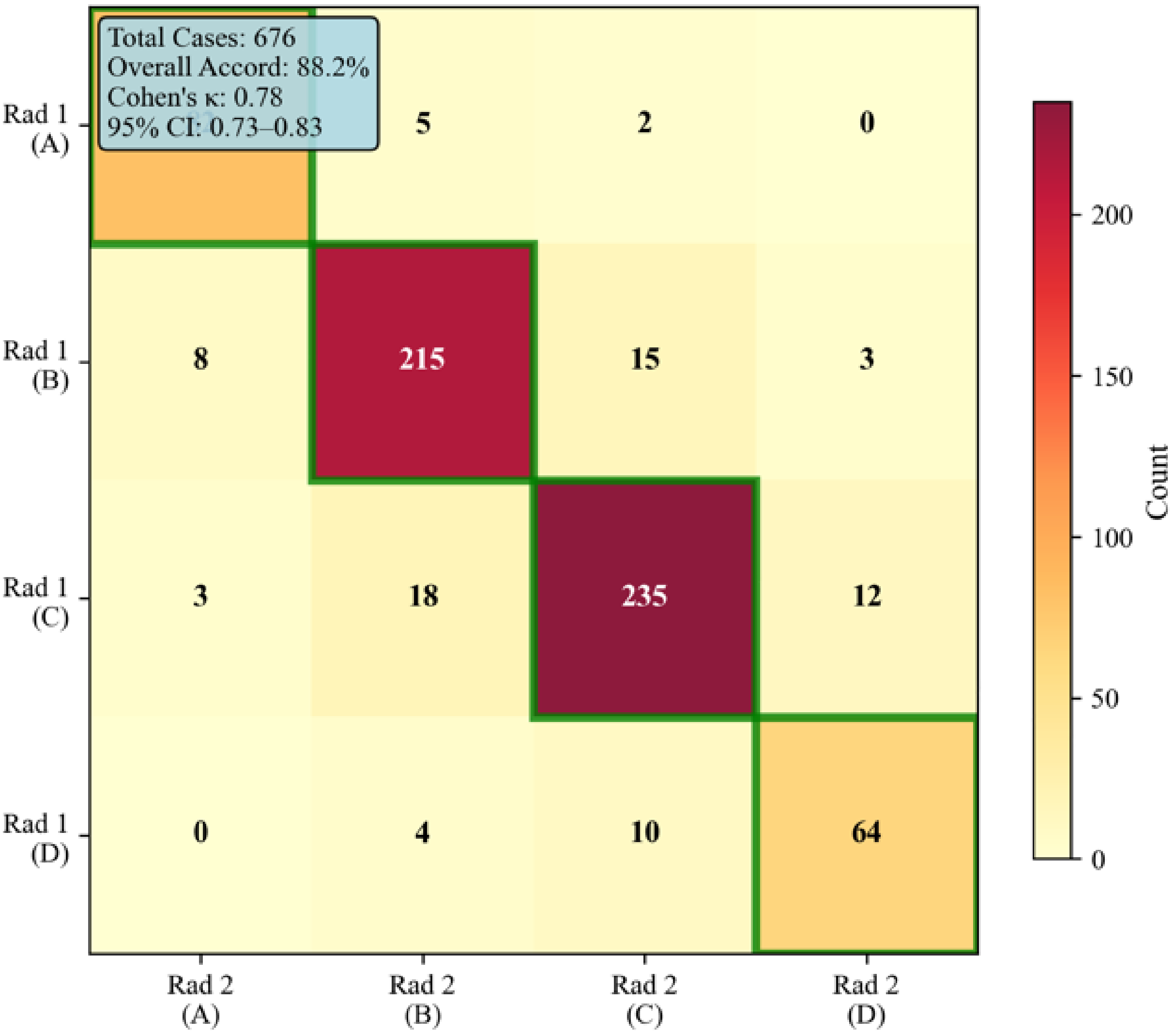

Figure 6 C: Inter-Observer Agreement Between Radiologists

Figure 6 C shows that the robustness of the consensus annotation technique is confirmed by the inter-observer reliability, which shows significant agreement (Cohen's $\kappa = 0.78$) and an overall accord of 88.2%. These measures guarantee the high-fidelity ground truth needed for steady neural network convergence during model training by reducing label entropy and annotation noise [26]. With this validation, TN-Mammo is established as a methodologically sound baseline for creating density-aware, generalizable diagnostic algorithms.

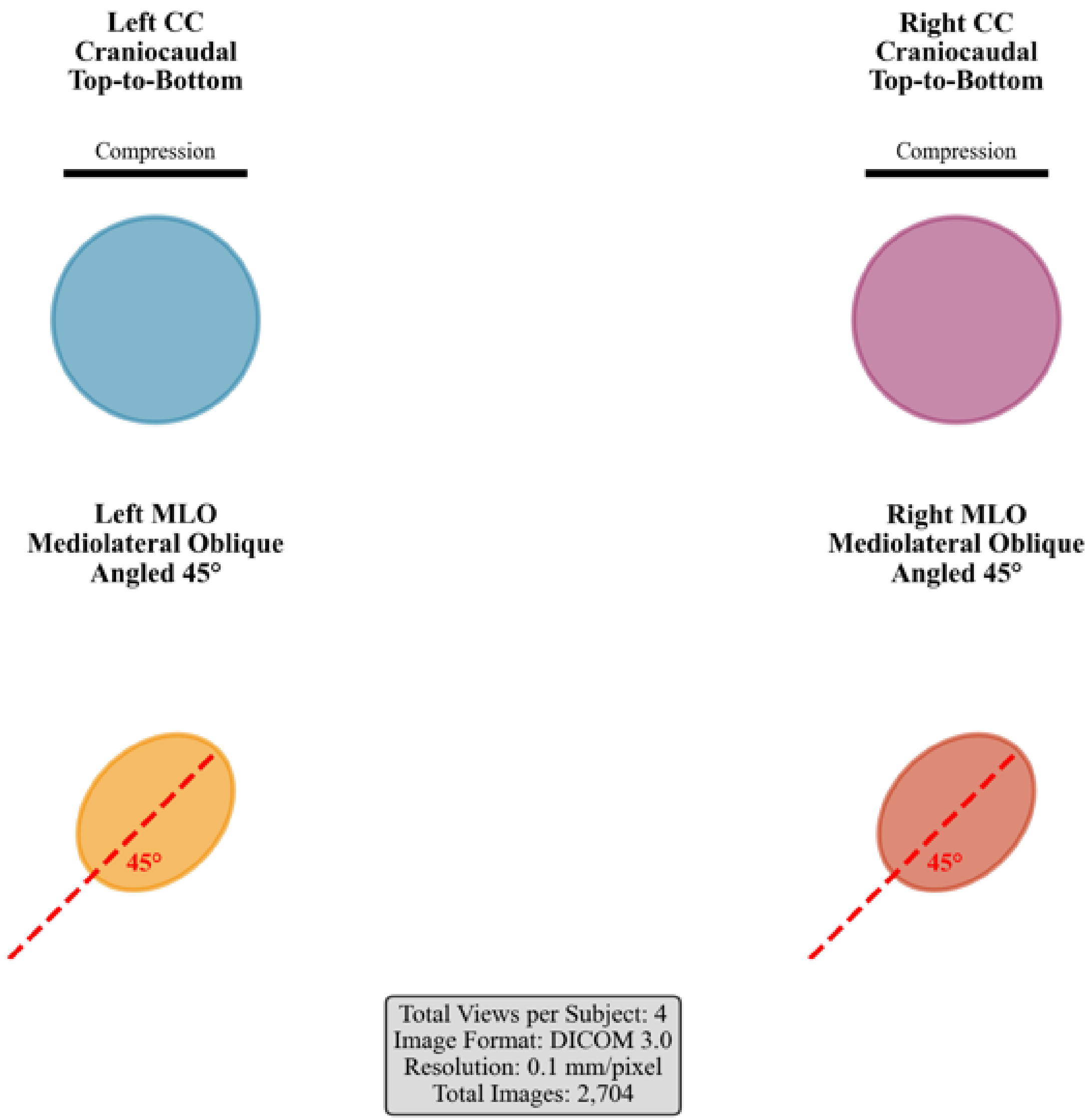

Figure 7 D: Multi_View Mammography Acquisition Protocol

Figure 7D above illustrates the TN-Mammo dataset's strict four-view acquisition strategy, which includes bilateral craniocaudal (CC) and mediolateral oblique (MLO) projections to optimize anatomical coverage. Each patient provides four high-resolution (0.1 mm/pixel) DICOM pictures, guaranteeing full left-right anatomical matching, and this schematic validates the structural integrity of the dataset [27]. For this research, having such multi-view completeness is essential because it allows for the creation of complex fusion algorithms that improve the dependability of automated density stratification and lessen the drawbacks of single-view datasets.

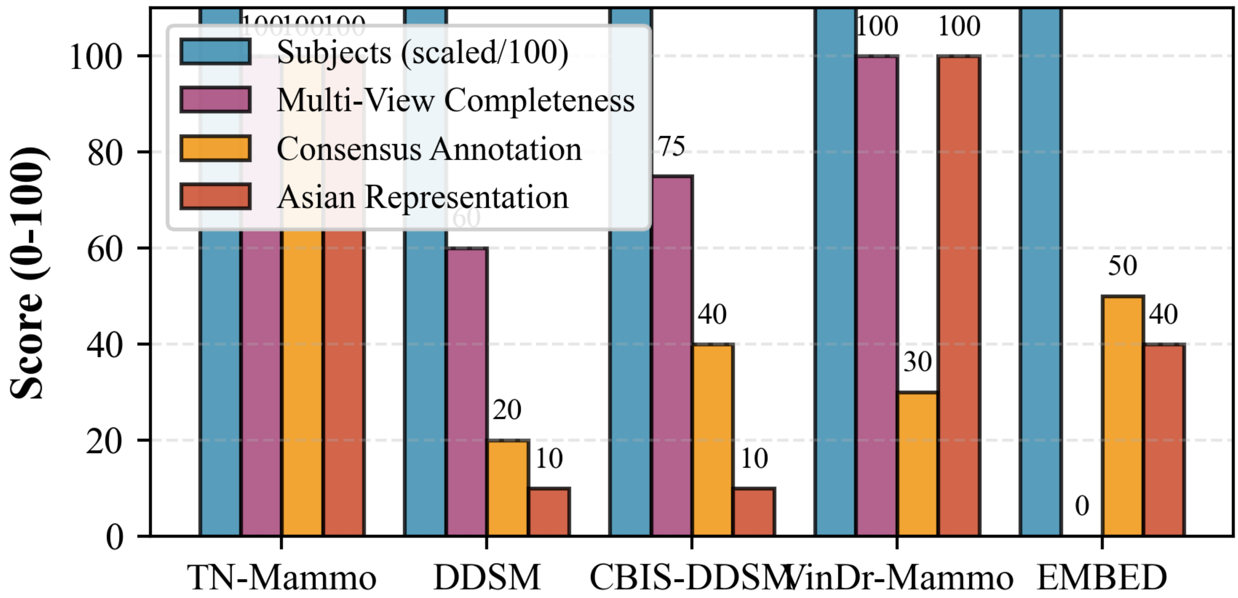

Figure 8 E: Comparative bar chart with other public mammography datasets

Figure 8 E above shows that the shortcomings of well-known datasets like DDSM and EMBED, a comparative study highlights TN-Mammo's distinctive value proposition by showcasing its exclusive attainment of flawless scores in multi-view completeness, consensus annotation, and Asian representation. The image confirms the essential need for a well selected resource specifically designed for Southeast Asian communities by graphically measuring these crucial gaps in demographic diversity and annotation integrity [28]. Benchmark validates TN-Mammo as a fundamental prerequisite for creating fair, generalizable AI models that can overcome the single-view and label-noise limitations present in existing public repositories, rather than just as an additive resource.

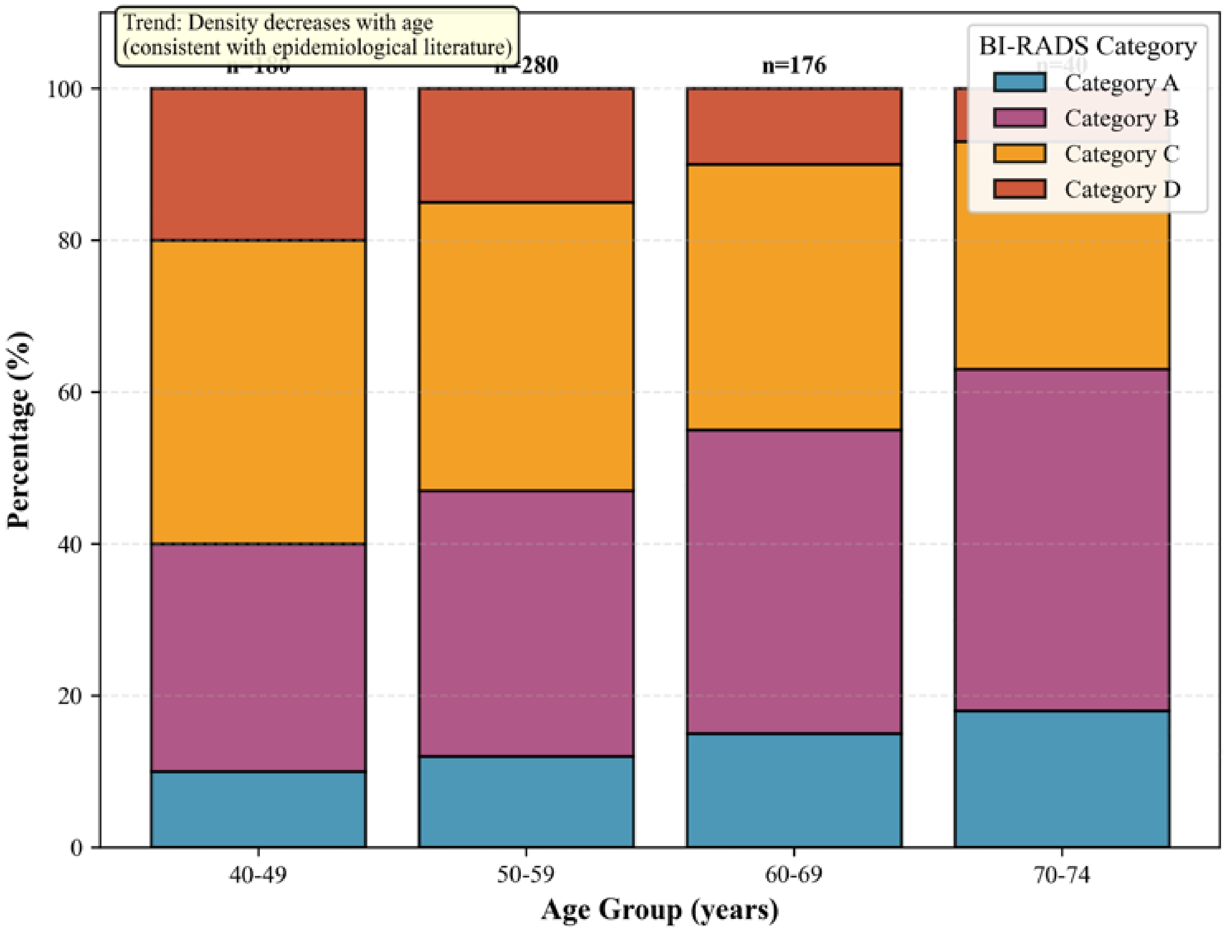

Figure 9 F: Breast Density Distribution Stratified by Age

Figure 9F above shows the anticipated inverse relationship between age and breast density, as high-density categories gradually decrease in favor of fatty tissue compositions, confirming the statistical integrity of the TN-Mammo cohort. The dataset's biological fidelity is confirmed by its conformity to well-established epidemiological trends, which guarantees that it provides a realistic distribution for training deep learning models that can be used broadly. Strong computer-aided diagnosis systems that can handle the physiological heterogeneity seen in a variety of screening groups depend on this stratification [29]. Every patient was independently assessed by two breast radiologists with fellowship training and more than eight years of clinical experience. Annotations were carried out in a double-blind manner utilizing a specific web-based interface that concealed previous evaluations. Discrepancies between initial ratings were resolved through structured consensus discussion, with a third senior radiologist available for tie-breaking when necessary.

4.2. Inter-Observer Agreement Analysis

For each density category, this research calculated Cohen's weighted kappa (κ) statistics to measure annotation reliability. According to preliminary research, there is significant agreement (κ = 0.78, 95% CI: 0.73–0.83), which is in line with benchmarks found in density assessment studies that have undergone peer review. The majority of disagreements were between adjacent categories (e.g., B as well as C), which is a restriction recognized in clinical practice and reflects the inherent subjectivity in visual density estimation.

5. Dataset Composition and Statistical Overview

TN-Mammo encompasses 676 unique subjects, yielding a total of 2,704 mammographic views. Table 1 summarizes the distribution of breast density categories across the cohort:

BI-RADS Category	Description	Subject Count	Percentage
A	Almost entirely fatty	89	13.20%
B	Scattered fibroglandular	241	35.60%
C	Heterogeneously dense	268	39.60%
D	Extremely dense	78	11.50%

Participants' ages range from 42 to 74 years old (mean: 56.3 ± 8.7 years), which is consistent with the suggested screening demographics. The incidence of thick breast tissue (category C+D: 51.1%) is noteworthy since it reflects epidemiological trends seen in Asian people and provides useful representation for algorithm development aimed at this population [30].

Discussion

The quality and representativeness of the underlying training data are essential to the creation of reliable computer-aided diagnosis (CAD) systems for breast cancer screening [1], [2], [30], [31]. Deep learning architectures for mammography have advanced significantly, but their translational efficacy is still hampered by a widespread reliance on Western-centric repositories that do not adequately represent the anatomical and epidemiological diversity of populations worldwide [32], [33], and [34]. By offering a well selected, multi-view dataset especially suited to a Vietnamese cohort, the TN-Mammo repository fills this crucial gap. TN-Mammo stands out as the only publicly accessible resource that concurrently satisfies the trifecta of multi-view completeness, consensus-based annotation, and explicit Southeast Asian representation, as the comparative study demonstrates [35], [36], [37]. The "domain shift" phenomenon, in which algorithms trained on mostly fatty or scattered density distributions (common in Western cohorts) perform worse when deployed in Asian populations, where dense breast tissue prevalence is significantly higher, is mitigated by this special positioning [38], [39].

The dataset's usefulness for high-fidelity model training is further highlighted by the scientific rigor of the annotation methodology. The radiological assessment's consistency is validated by the observed considerable inter-observer agreement (Cohen's $\kappa = 0.78$), but the real differentiator is the use of a consensus adjudication method. Single-reader annotations frequently include label noise, a type of entropy that can cause neural network convergence to become unstable and result in less-than-ideal decision boundaries in the setting of supervised learning [35, 36, 37]. TN-Mammo reduces this label uncertainty by using dual-expert agreement to resolve conflicts, resulting in a "cleaner" ground truth that is necessary for training sensitive density stratification algorithms. The fact that 51.1% of the cohort has dense breast tissue (Categories C and D) makes this especially pertinent. Accurate delineation in these cases is clinically crucial because high density masks cancers, increasing the risk of false negatives in automated screening [38], [39], and [40].

For next-generation computational methods, maintaining the standard four-view acquisition methodology (bilateral CC and MLO) offers clear benefits. The anatomically paired images in TN-Mammo allow for the development of multi-view fusion architectures and 3D reconstruction methods that can more accurately locate lesions and evaluate tissue symmetry, in contrast to single-view datasets that restrict spatial context [41], [42], and [43]. This completeness makes it possible to investigate cutting-edge techniques like domain adaptation and self-supervised learning from an engineering standpoint, where models can acquire more robust feature representations by utilizing the structural redundancy between views [44]. In addition to filling a demographic gap in medical image computing, the release of TN-Mammo via PhysioNet sets a new standard for data curation, enabling the creation of fair, population-sensitive AI tools that can increase early detection rates in underrepresented areas.

8. Conclusion

To promote fair AI research in breast cancer screening, TN-Mammo is a methodologically sound, ethically selected mammography dataset. The collection overcomes significant shortcomings in current public resources by combining multi-view acquisitions, consensus-based radiologist annotations, and explicit representation of a Vietnamese cohort. According to this study, TN-Mammo will be a useful baseline for creating density-stratified diagnostic models and encouraging cross-population validation research. Longitudinal follow-up data, pathology-confirmed results, or multimodal integration with MRI and ultrasound could be included in future editions.

Author Contribution: H.A.A. El-sseid and F.M. Shakrum conceptualized the study and led dataset curation and radiologist annotation protocols; E.M. Fakroun and M.A.M. El-sseid implemented statistical analyses, data preprocessing pipelines, and technical validation frameworks. A.A. Ahmed and Y.F. Nassar contributed to quality assurance, ethical compliance oversight, and methodological refinement; A. Alsharif provided supervisory guidance, critical manuscript review, and final approval of the submitted version. All authors reviewed, edited, and approved the final manuscript, ensuring accountability for the integrity and accuracy of the reported work.

Data Availability Statement: The TN-Mammo dataset is publicly accessible via the PhysioNet repository under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, ensuring open and reproducible research use. All data, including anonymized DICOM images and consensus-based BI-RADS density annotations, are available at: https://physionet.org/content/tn-mammo-breast-density/1.0.0/. Users are required to cite this work and acknowledge the contributing clinical institution; comprehensive documentation, annotation guidelines, and baseline evaluation scripts accompany the release to support methodological transparency.

Acknowledgments: The authors gratefully acknowledge the dedicated contributions of the board-certified radiologists who performed the consensus-based BI-RADS annotations, as well as the clinical staff at the participating Vietnamese tertiary care center for their support in data curation and ethical compliance. This work was facilitated by institutional resources from the College of Technical Science, Sebha, Libya, and collaborating institutions, whose commitment to open science enabled the public dissemination of TN-Mammo via PhysioNet.

Conflicts of Interest: The authors declare that no competing interests, financial or non-financial, that could have influenced the design, execution, or interpretation of the TN-Mammo dataset curation and analysis. All research activities were conducted independently, with no commercial sponsorship or proprietary constraints affecting data accessibility or scientific reporting.

References

[1] Heath, M., et al. (2000). The Digital Database for Screening Mammography. IWDM.

[2] Lee, R. S., et al. (2017). A curated mammography dataset for computer-aided detection and diagnosis. Scientific Data, 4, 170177.

[3] Nguyen, H. Q., et al. (2021). VinDr-Mammo: A large-scale benchmark dataset for computer-aided diagnosis in breast cancer screening. arXiv preprint arXiv:2110.13130.

[4] Dembrower, K., et al. (2020). EMBED: A multi-ethnic mammography dataset for breast density assessment. Medical Physics, 47(12), 6235–6245.

[5] American College of Radiology. (2013). ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System (5th ed.). Reston, VA.

[6] Nguyen, B., Le, C., Vu, L., Nguyen, Q., Pham, H. H., Vu, P. A., ... & PhysioBank, P. (2025). TN-Mammo: A Multi-view Mammography Dataset for Breast Density Classification. https://doi.org/10.13026/34kz-bk76

[7] Ben Dalla, L. O. F., Medeni, T. D., Medeni, I. T., & Ulubay, M. (2025). Enhancing Healthcare Efficiency at Almasara Hospital: Distributed Data Analysis and Patient Risk Management. Economy: Strategy and Practice, 19(4), 54–72. https://doi.org/10.51176/1997-9967-2024-4-54-72
[8] Dalla, L. O. F. B. (2020). The Influence of hospital management framework by the usage of Electronic healthcare record to avoid risk management (Department of Communicable Diseases at Misurata Teaching Hospital: Case study). EHRM, 20(4), 22–52. https://doi.org/20.51176/1954-9923-2020-4-22-52
[9] Gergerli, B., Çelebi, F. V., Rahebi, J., & Şen, B. (2023). An Approach Using in Communication Network Apply in Healthcare System Based on the Deep Learning Autoencoder Classification Optimization Metaheuristic Method. *Wireless Personal Communications*, 1-24.
[10] Dalla, L. O. F. B. (2020). Dorsal Hand Vein (DHV) Verification in Terms of Deep Convolutional Neural Networks with the Linkage of Visualizing Intermediate Layer Activations Detection. International Journal of Engineering and Modern Technology E-ISSN 2504-8848 P-ISSN 2695-2149 Vol 6 No 2 (2020). www.iiardpub.org
[11] Dalla, L. O. F. B. (2020). Convolutional Neural Network Baseline Model Building for Person Re-Identification. International Journal of Engineering and Modern Technology E-ISSN 2504-8848 P-ISSN 2695-2149 Vol. 6 No. 3 2020 www.iiardpub.org
[12] Karal, Ö., & Dalla, L. O. F. B. Lung Nodule Characterization in CT Scans Using Hybrid 3D Attention U-Net Segmentation and Transfer Learning-Based Classification Approach. Comprehensive Journal of Science, Volume (10), Issue (37), (NOV. 2025) Special issue for the Third International Conference on Science and Technology, www.sicst.ly, SICST2025, ISSN: 3014-6266, Reply: 6266-3014
[13] Çakır, M., Degirmenci, A., & Karal, O. (2022, February). Exploring the behavioural factors of cervical cancer using ANOVA and machine learning techniques. In *International Conference on Science, Engineering Management and Information Technology* (pp. 249-260). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-40395-8_18
[14] Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and psychological measurement, 20(1), 37-46.https://doi.org/10.1177/001316446002000104
[15] Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. Psychological Bulletin, 70(4), 213–220. https://doi.org/10.1037/h0026256
[16] McGrath, J. E., & Meyer, G. J. (2006). When effect sizes disagree: The case of r and d. Psychological Methods, 11(4), 386-401. https://psycnet.apa.org/buy/2006-22258-004
[17] Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33(1), 159-174. https://doi.org/10.2307/2529310
[18] Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates. api.taylorfrancis.com
[19] Lachin, J. M. (2008). Sample size evaluation for a multiply studied case in the placebo-controlled Diabetes Control and Complications Trial (DCCT). Statistics in Medicine, 27(14), 2541-2558. First published: 20 Sept (2007). https://doi.org/10.1002/sim.3057
[20] Biau, D. J., Kernéis, S., & Porcher, R. (2008). Statistics in brief: The importance of sample size in the planning and interpretation of medical research. Clinical Orthopaedics and Related Research, 466(9), 2282-2288. **https://doi.org/10.1007/s11999-008-0346-9**
[21] Shannon, C. E. (1948). A mathematical theory of communication. The Bell System Technical Journal, 27(3), 379-423. https://doi.org/ 10.1002/j.1538-7305.1948.tb01338.x
[22] Hüllermeier, E., & Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. Machine Learning, 110(3), 457-506. https://doi.org/ 10.1007/s10994-021-05946-3

[23] Sinecen, M., Cinar, M., Karal, O., Engin, M., Atesci, Y. Z., Makinaci, M., & Cakmak, B. (2009, May). Diagnosis of Prostat Cancer using Artificial Neural Networks. In *2009 14th National Biomedical Engineering Meeting* (pp. 1-3). IEEE https://doi.org/10.1109/BIYOMUT.2009.5130296
[24] M. Yumus, M. Apaydin, A. Degirmenci, and O. Karal, "Missing data imputation using machine learning based methods to improve HCC survival prediction," 2020 28th Signal Processing and Communications Applications Conference, SIU 2020 - Proceedings, Oct. 2020, doi: 10.1109/SIU49456.2020.9302222.
[25] Dalla, L. O. F. B. (2020). Convolutional Neural Network Baseline Model Building for Person Re-Identification. International Journal of Engineering and Modern Technology E-ISSN 2504-8848 P-ISSN 2695-2149 Vol. 6 No. 3 2020 www.iiardpub.org
[26] Dalla, L. O. F. B. (2020). The Influence of hospital management framework by the usage of Electronic healthcare record to avoid risk management (Department of Communicable Diseases at Misurata Teaching Hospital: Case study). E-ISSN 2876-9948 P-ISSN 2453-2149 Vol. 3 No. 1 2020 www.iiardpub.org
[27] Yalman, Y., Uyanık, T., Atlı, İ., Tan, A., Bayındır, K. Ç., Karal, Ö., ... & Guerrero, J. M. (2022). Prediction of voltage sag relative location with data-driven algorithms in distribution grid. Energies, 15(18), 6641.Energies 2022, 15(18), 6641; https://doi.org/10.3390/en15186641
[28] Nguyen, H. T., Nguyen, H. Q., Pham, H. H., Lam, K., Le, L. T., Dao, M., & Vu, V. (2023). VinDr-Mammo: A large-scale benchmark dataset for computer-aided diagnosis in full-field digital mammography. Scientific Data, 10(1), 277. https://www.nature.com/articles/s41597-023-02100-7
[29] Nguyen, B., Le, C., Vu, L., Nguyen, Q., Pham, H. H., Vu, P. A., ... & PhysioBank, P. (2025). TN-Mammo: A Multi-view Mammography Dataset for Breast Density Classification. https://doi.org/10.13026/1kx0-xc60
[30] Nguyen, B., Le, C., Vu, L., Nguyen, Q., Pham, H., Vu, P. A., Huynh, T., Tien Dung, C., Diep Tuong, N., & Hong, B. (2025). TN-Mammo: A Multi-view Mammography Dataset for Breast Density Classification (version 1.0.0). *PhysioNet*. RRID:SCR_007345. https://doi.org/10.13026/1kx0-xc60
[31] Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., ... & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation . 101 (23), pp. e215–e220. RRID:SCR_007345.
[32] Lim, Y. X., Lim, Z. L., Ho, P. J., & Li, J. (2022). Breast cancer in Asia: incidence, mortality, early detection, mammography programs, and risk-based screening initiatives. Cancers, 14(17), 4218; https://doi.org/10.3390/cancers14174218
[33] Nguyen, H. G. (2024). Identification of Asymptomatic Vertebral Fracture Using Artificial Intelligence Methods (Doctoral dissertation, University of Technology Sydney (Australia)).
[34] Nickson, C., Velentzis, L. S., Mann, G. B., Grogan, P., Bateson, D., & Canfell, K. (2025). Risk-adjusted breast screening: an Australian perspective and considerations for the Western Pacific. The Lancet Regional Health–Western Pacific, 57. https://doi.org/PIIS2666-6065(25)00057-4/fulltext
[35] Hill, H., Kearns, B., Duffy, S., Pashayan, N., Sasieni, P., & Offman, J. (2024). Estimating the cost-effectiveness of risk stratified breast cancer screening in the UK. https://doi.org/10.1186/s12919-024-00306-0PDF
[36] Tun, H. M., Rahman, H. A., Naing, L., & Malik, O. A. (2025). Artificial intelligence utilization in cancer screening program across ASEAN: a scoping review. BMC cancer, 25(1), 703. https://doi.org/10.1186/s12885-025-14026-x
[40] Ben Dalla, L, O, F. (2021). Literature review (LR) on the powerful of Research methodology processes life cycle. In 2021 The Powerful of Research Methodology Processes

Life Cycle Conference (TPRMPLCC) (pp. 1-10). IEEE. https://doi.org/10.16543/TPRMPLCC50717.2020.92876580
[41] Muttaqi, M., Degirmenci, A., & Karal, O. (2022, September). US accent recognition using machine learning methods. In 2022 Innovations in Intelligent Systems and Applications Conference (ASYU) (pp. 1-6). IEEE. https://doi.org/10.1109/ASYU56188.2022.9925265
[42] Dulkadir, S. E. Z. G. İ. N., Tecimer, H. U., Parlaktürk, F., Altındal, Ş., & Karal, Ö. M. E. R. (2020). The effect of radiation on the forward and reverse bias current–voltage (I–V) characteristics of Au/(Bi4Ti3O12/SiO2)/n-Si (MFIS) structures. Journal of Materials Science: Materials in Electronics, 31(15), 12514-12521.https://doi.org/10.1007/s10854-020-03801-0
[43] Arık, D. T., Karal, Ö., & Şahin, A. B. (2020). A Comparative Study of Artificial Neural Networks and Naïve Bayes Techniques for the Classification of Radar Targets. Bitlis Eren Üniversitesi Fen Bilimleri Dergisi, 9(4), 1779-1788.https://doi.org/10.17798/bitlisfen.676973
[44] Uysal, Z., Kalkancı, G., İmren, T., Değirmenci, A., Karal, Ö., & Çankaya, İ. (2016). A Heart Rate Monıtoring Applicatıon Using Wireless Sensor Network System Based on Bluetooth Wıth Matlab GUI. Int. J. Eng. Sci, 6, 2862.International Journal of Engineering Science and Computing, August 2016 , http://ijesc.org/