

ORIGINAL

Estudio de la variabilidad inter- e intraobservador en la determinación de la densidad mamaria según la 5.^a edición del Atlas BI-RADS®



K. Pesce^a, M. Tajerian^b, M.J. Chico^a, M.P. Swiecicki^a, B. Boietti^b, M.J. Frangella^{b,*} y S. Benitez^b

^a Servicio de Diagnóstico por imágenes, Hospital Italiano de Buenos Aires, Buenos Aires, Argentina

^b Departamento de Informática en Salud, Hospital Italiano de Buenos Aires, Buenos Aires, Argentina

Recibido el 22 de octubre de 2019; aceptado el 15 de abril de 2020

Disponible en Internet el 31 de mayo de 2020

PALABRAS CLAVE

Densidad mamaria;
Variabilidad
intraobservador;
Variabilidad
interobservador;
BI-RADS® 5 edición;
Densidad
mamográfica

Resumen

Objetivo: Determinar el acuerdo intra- e interobservador en la categorización de la densidad mamográfica entre un grupo de profesionales según la 5.^a edición del Atlas BI-RADS® - ACR y analizar la concordancia entre la categorización de los expertos y un *software* comercial de un mamógrafo digital para categorización automática.

Métodos: 6 médicos categorizaron la densidad mamográfica de 451 mamografías en dos oportunidades con un intervalo de 1 mes. Calculamos los coeficientes kappa ponderados lineales de acuerdo inter- e intraobservador para el grupo médico y la concordancia entre el *software* comercial y el reporte de la mayoría. Analizamos los resultados para las cuatro categorías de densidad mamaria y para el resultado dicotómico de mama densa/no densa.

Resultados: El acuerdo interobservador entre especialistas y el reporte de la mayoría fue moderado y casi perfecto para el análisis por categoría ($\kappa = 0,64$ a $0,84$) y de manera dicotómica ($\kappa = 0,63$ a $0,84$). El acuerdo intraobservador fue sustancial y casi perfecto ($\kappa = 0,68$ a $0,85$ para 4 categorías y $\kappa = 0,70$ a $0,87$ para el análisis dicotómico). El acuerdo entre el reporte de la mayoría y el *software* comercial fue moderado tanto por categoría ($\kappa = 0,43$) como en el análisis dicotómico ($\kappa = 0,51$).

Conclusión: Hemos observado un acuerdo entre moderado y casi perfecto inter- e intraobservador entre los radiólogos, según los criterios establecidos en la 5.^a edición del Atlas BI-RADS®. El nivel de acuerdo entre el reporte de los especialistas y un *software* disponible comercialmente fue moderado.

© 2020 SERAM. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

* Autor para correspondencia.

Correo electrónico: maria.frangella@hospitalitaliano.org.ar (M.J. Frangella).

KEYWORDS

Breast density;
Intraobserver
variability;
Interobserver
variability;
BI-RADS® fifth
edition;
Breast density on
mammograms

Interobserver and intraobserver variability in determining breast density according to the fifth edition of the BI-RADS® Atlas

Abstract

Objective: To determine the level of agreement within and between observers in the categorization of breast density on mammograms in a group of professionals using the fifth edition of the American College of Radiology's BI-RADS® Atlas and to analyze the concordance between experts' categorization and automatic categorization by commercial software on digital mammograms.

Methods: Six radiologists categorized breast density on 451 mammograms on two occasions one month apart. We calculated the linear weighted kappa coefficients for inter- and intra-observer agreement for the group of radiologists and between the commercial software and the majority report. We analyzed the results for the four categories of breast density and for dichotomous classification as dense versus not dense.

Results: The interobserver agreement among radiologists and the majority report was between moderate and nearly perfect for the analysis by category ($\kappa = 0.64$ to 0.84) and for the dichotomous classification ($\kappa = 0.63$ to 0.84). The intraobserver agreement was between substantial and nearly perfect ($\kappa = 0.68$ to 0.85 for 4 categories and $\kappa = 0.70$ to 0.87 for the dichotomous classification). The agreement between the majority report and the commercial software was moderate both for the four categories ($\kappa = 0.43$) and for the dichotomous classification ($\kappa = 0.51$).

Conclusion: Agreement on breast density within and between radiologists using the criteria established in the fifth edition of the BI-RADS® Atlas was between moderate and nearly perfect. The level of agreement between the specialists and the commercial software was moderate.

© 2020 SERAM. Published by Elsevier España, S.L.U. All rights reserved.

Introducción

La densidad mamaria, evaluada mediante mamografía, se define como las cantidades relativas de tejido estromal y epitelial radiodensos respecto al tejido adiposo radiotransparente¹. En la 5.ª edición del Atlas BI-RADS® del American College of Radiology (ACR) se definen cuatro patrones de densidad mamaria: ACR a predominantemente adiposo, ACR b con regiones dispersas de tejido fibroglandular, ACR c heterogéneamente denso y ACR d extremadamente denso².

La alta densidad mamaria (categorías c y d) es un factor de riesgo independiente para el desarrollo del cáncer de mama³⁻⁵ y un factor de enmascaramiento que conlleva una disminución de la sensibilidad de la mamografía para la detección de esta enfermedad y otras lesiones⁶⁻¹⁰.

La evaluación de la densidad mamaria puede ser realizada de forma visual por un radiólogo. Algunos autores han detectado que este método supone una variabilidad intra- e interobservador considerable¹¹⁻¹³. En una revisión sistemática de la bibliografía, realizada por el grupo de trabajo Task Force en el año 2016, se reportó que, en entornos comunitarios, el 19-22% de los exámenes mamográficos inicialmente clasificados como densos se reclasificaron posteriormente como no densos, mientras que el 10-16% de los exámenes inicialmente no densos se reclasificaron como densos. En la reproducibilidad secuencial describieron que al 20% de los estudios se le modifica la categoría en la ronda subsiguiente si la lectura era realizada por el mismo radiólogo, y la proporción subía al 33% si era evaluado por otro radiólogo^{13,14}.

Por otra parte, cada vez son más frecuentes los sistemas automatizados de medición de la densidad mamaria¹⁵⁻¹⁹. Consecuentemente, en la práctica clínica, la densidad mamaria se reporta a partir de la valoración que realizan los especialistas con el soporte de dichos sistemas. Pero son escasos los estudios que realizan un análisis acerca del uso, la concordancia y la percepción de utilidad de los profesionales sobre estas herramientas.

El objetivo de nuestro trabajo fue determinar el acuerdo intra- e interobservador en la categorización de la densidad mamográfica según la 5.ª edición del Atlas BI-RADS® - ACR, entre un grupo de profesionales especializados en imágenes mamarias de una institución de salud de alta complejidad. Analizamos además la concordancia entre la categorización realizada por expertos y la categorización realizada por un método de categorización automático: *software* comercial de un mamógrafo digital.

Métodos**Ámbito**

Nuestro estudio fue llevado a cabo en la Sección de Diagnóstico e Intervencionismo Mamario, del Servicio de Diagnóstico por Imágenes de un hospital de tercer nivel. El servicio cuenta con imágenes digitales y sistema integrado RIS/PACS desde el año 2010²⁰. La sección está compuesta por 10 especialistas, 2 *fellows* e informa un promedio de 30.000 mamografías anuales. Los estudios mamográficos se asignan diariamente a los médicos radiólogos al azar para su reporte,

cada uno recibe entre 200 y 400 casos mensuales. Una vez redactados los reportes de mamografía, se somete a revisión de pares el 10% de los estudios informados por especialistas (aproximadamente 300 estudios mensuales) y la totalidad de los estudios informados por *fellows*. Además, se realizan auditorías de calidad de reporte por parte del médico que solicitó el estudio.

Diseño del estudio

Este estudio de corte transversal se realizó de acuerdo con los principios de la Declaración de Helsinki y fue aprobado por el Comité de Ética de nuestra institución. Contó, además, con el consentimiento de los pacientes.

Categorización de la densidad mamaria según la 5.ª edición del BI-RADS® -- ACR

El equipo de profesionales estuvo conformado por 6 médicos de la Sección de Diagnóstico e Intervencionismo Mamario. El grupo contaba con un promedio de experiencia en imágenes mamarias de 9 años y un rango de 2 a 18 años. Se incluyeron un total de 451 mamografías de pacientes asintomáticas seleccionadas al azar, perteneciente a pacientes con un rango etario entre 40 y 90 años, realizadas en la institución durante el mes de febrero de 2019. Se extrajo una de las cuatro adquisiciones de cada estudio, que podían ser craneocaudales u oblicuomediolaterales. Se excluyeron las incidencias focalizadas, magnificadas y estudios mamográficos de pacientes con antecedentes personales de cirugía mamaria (incluyendo implantes mamarios) y gigantomastia. Se definió esta última según la necesidad de utilizar más de una placa por incidencia.

Un bioingeniero y un radiólogo, que no participaron en la posterior categorización, extrajeron las imágenes para evaluar de la base de datos del hospital y eliminaron los datos de identificación de las pacientes. El muestreo utilizado fue de tipo consecutivo.

Una semana antes de la evaluación de las 451 imágenes mamográficas los profesionales participantes repasaron los criterios de categorización de densidad mamaria con imágenes de la 5.ª edición del Atlas BI-RADS®². A continuación, los especialistas categorizaron la densidad mamográfica de cada una de las mamografías. Las imágenes fueron evaluadas en estaciones de trabajo de 5 megapíxeles.

Los evaluadores no tenían conocimiento de los datos demográficos de las pacientes ni la categoría asignada en el reporte original de cada mamografía. Tampoco conocían la densidad asignada por los demás participantes del estudio ni la evaluación del *software* comercial. Se realizaron dos lecturas de las mismas mamografías separadas con un intervalo de un mes. El orden de los estudios en cada una de las dos lecturas fue aleatorio. Toda esta información se registró en una base de datos electrónica.

Utilizamos un *software* comercialmente disponible para categorización automática de densidad mamaria de las imágenes de nuestra muestra: Breast Density Measurement Software, del mamógrafo AMULET Innovality –3000AWS7.0 Option– FUJIFilm®. Este *software* calcula la densidad mamaria a partir de la relación entre el tejido fibroglandular y el área total de la mama²¹. A partir de esta propor-

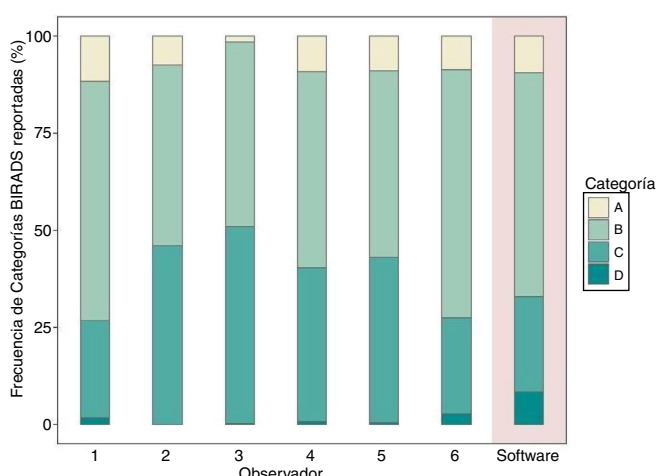


Figura 1 Distribución de frecuencias de categorías de densidad mamaria, por especialista.

ción, estima la densidad mamaria en una escala continua y automáticamente asigna una categoría.

Variables de interés y métodos estadísticos

Evaluamos la variabilidad intra- e interobservador de cada uno de los profesionales participantes mediante el cálculo de coeficientes kappa ponderados lineales, que se reportan con un intervalo de confianza (IC) del 95%. El cálculo del coeficiente se realizó para las cuatro categorías de densidad mamaria y también para el resultado dicotómico de mama densa versus mama no densa.

Elegimos, además, la medida de reporte de la mayoría (*majority report*), definida como la moda estadística de los reportes de los observadores, en concordancia con la literatura preexistente²². Calculamos el nivel de acuerdo entre cada evaluador y el reporte de la mayoría. Para los casos en los que se produjo un empate (distribución de categorizaciones no unimodal), un séptimo especialista en imágenes categorizó la mamografía para alcanzar un acuerdo.

Para el cálculo de kappa ponderado lineal (κ) utilizamos el método descrito por Cohen y Fleiss^{23,24}. Reportamos los coeficientes kappa de acuerdo interobservador e intraobservador, que fueron a su vez clasificados según su nivel de acuerdo siguiendo la subdivisión del coeficiente kappa propuesta por Landis y Koch (0: "pobre"; de 0 a 0,2: "leve"; de 0,21 a 0,4: "justo"; de 0,41 a 0,6: "moderado"; de 0,61 a 0,8: "sustancial"; de 0,81 a 1: "casi perfecto")²⁵.

Por último, reportamos la concordancia entre el *software* comercial y el reporte de la mayoría.

Utilizamos el *software* estadístico STATA v. 14 y R versión 3.6.0. Se consideró un valor de $p < 0,05$ para indicar significación.

Los investigadores participantes declaran no tener conflictos de intereses.

Resultados

La frecuencia de las categorías asignadas por cada especialista se muestra en la [figura 1](#). Los valores de kappa ponderados lineales para el acuerdo global interobservador

Tabla 1 Kappa ponderado lineal entre cada especialista y el reporte de mayoría

	Kappa ponderado lineal (IC 95%)	
	Interobservador En escala de 4 categorías ^a	Interobservador Dicotomizado (denso/no denso) ^b
Observador 1	0,66 (0,60 a 0,72)	0,63 (0,55 a 0,70)
Observador 2	0,77 (0,72 a 0,83)	0,76 (0,70 a 0,82)
Observador 3	0,64 (0,58 a 0,70)	0,72 (0,66 a 0,78)
Observador 4	0,84 (0,80 a 0,89)	0,84 (0,79 a 0,89)
Observador 5	0,83 (0,78 a 0,87)	0,80 (0,75 a 0,86)
Observador 6	0,67 (0,61 a 0,73)	0,66 (0,59 a 0,73)
Software comercial	0,46 (0,39 a 0,52)	0,51 (0,43 a 0,59)

^a Acuerdo interobservador para la primera etapa de observación en las 4 categorías de densidad BI-RADS®.

^b Acuerdo interobservador para la primera etapa de observación para el resultado dicotomizado (mama densa/no densa).

(entre cada especialista y el reporte de la mayoría) en el momento de la primera evaluación se muestran en la [tabla 1](#). Estos valores alcanzaron un rango de 0,64 (IC: 0,58 a 0,70) a 0,84 (IC: 0,80 a 0,89) para el análisis por categoría. Es decir, se alcanzó un acuerdo de moderado a casi perfecto entre cada uno de los observadores y el reporte de la mayoría. Valores similares se obtuvieron al analizar los resultados de manera dicotómica desde una perspectiva clínica, categorizando la imagen como mama densa/no densa.

Los valores de kappa ponderados lineales para el acuerdo intraobservador entre la primera y la segunda observación se muestran en la [tabla 2](#). Los resultados alcanzaron un rango de 0,68 (IC: 0,63 a 0,74) a 0,85 (IC: 0,80 a 0,89). Para el análisis dicotómico, los resultados fueron de 0,70 (IC: 0,64 a 0,77) a 0,87 (IC: 0,83 a 0,92). Para los resultados antes mencionados, los valores corresponden a un acuerdo entre sustancial y casi perfecto.

El acuerdo entre el reporte de la mayoría y el *software* comercial fue moderado para las cuatro categorías y para el análisis dicotómico, con valores kappa de 0,46 (IC: 0,39 a 0,52) y 0,51 (IC: 0,43 a 0,59), respectivamente.

Discusión

La densidad mamaria es un factor de riesgo independiente para el desarrollo de cáncer de mama. La categorización de la densidad mamaria resulta fundamental para realizar una evaluación de riesgo personalizada y complementar de manera eficiente los cribados poblacionales con estudios de mayor sensibilidad en pacientes con patrones densos, como la resonancia magnética o la ecografía mamaria²⁶. Dicho factor puede ser utilizado para la selección del método más adecuado de diagnóstico para cada paciente. Esto permite evitar las sobreprestaciones o subprestaciones de estudios complementarios, logrando, así, evitar las demoras en el acceso a estudios correspondiente.

El presente trabajo evaluó la variabilidad inter- e intraobservador existente en un servicio de diagnóstico por imágenes para la categorización de densidad mamaria, basada en la 5.ª ed. del Atlas BI-RADS®.

El acuerdo interobservador para la densidad mamaria fue sustancial en la concordancia entre los 6 radiólogos y el reporte de la mayoría, tanto para la escala de 4 categorías como para la categorización dicotómica (denso/no

denso). Analizando trabajos similares en la literatura científica encontramos resultados muy variables. Existen grupos que reportan mayor heterogeneidad en sus observaciones, con estudios que muestran concordancia interobservador entre los pares de radiólogos que van de leve a sustancial, con valores de kappa que varían de 0,02 a 0,72 (media = 0,46, IC 95%: 0,36 a 0,55)²⁷. Otros trabajos, en cambio, reportan valores que son similares a los que presentamos aquí^{22,27-30}. El hecho de que el grupo de médicos participantes del estudio cuente con actividades académicas periódicas, ateneos y actualizaciones, auditorías de reporte, etc. podría explicar este resultado.

Para la evaluación de la variabilidad intraobservador, nuestro estudio registró valores kappa con un acuerdo sustancial y casi perfecto³¹, tanto en una escala de 4 categorías como en la escala dicotómica, sin una diferencia estadísticamente significativa entre ellos. No obstante, debemos mencionar el potencial riesgo de sesgo por el breve lapso transcurrido entre las dos lecturas.

En relación con lo anterior, si bien encontramos trabajos que reportan una diferencia estadísticamente significativa para los observadores con una experiencia mayor de 10 años en el diagnóstico de estudios mamográficos²⁸, nuestros resultados no reflejaron lo mismo. Esto se puede explicar porque el equipo médico forma parte de una sección especializada con criterios diagnósticos homogéneos, reuniones de actualización periódicas y dedicación exclusiva al diagnóstico mamario. Como ya se ha mencionado, los evaluadores participaron de una sesión de actualización sobre criterios de categorización de densidad mamaria con anterioridad al inicio del estudio. Por lo tanto, nuestros resultados podrían generalizarse en centros con las mismas características.

Las principales ventajas de una herramienta de diagnóstico automatizada son su consistencia en el tiempo y su falta de variabilidad³². Es por esto por lo que un sistema automatizado independiente del observador permite mediciones reproducibles y debería ser más adecuado para una evaluación confiable y estandarizada. En este estudio, el acuerdo entre el reporte de la mayoría de los médicos y el *software* disponible comercialmente fue moderado. Podríamos atribuirlo, en primer lugar, a que los criterios asociados a los dos métodos de categorización son distintos. La categorización visual se fundamenta principalmente en

Tabla 2 Kappa ponderado lineal para cada especialista en las lecturas realizadas con un intervalo de tiempo de un mes

	Kappa ponderado lineal (IC 95%)	
	Intraobservador En escala de 4 categorías ^a	Intraobservador Dicotomizado (denso/no denso) ^b
Observador 1	0,76 (0,71 a 0,81)	0,75 (0,68 a 0,82)
Observador 2	0,70 (0,64 a 0,76)	0,70 (0,64 a 0,77)
Observador 3	0,85 (0,80 a 0,89)	0,87 (0,83 a 0,92)
Observador 4	0,72 (0,66 a 0,77)	0,72 (0,66 a 0,79)
Observador 5	0,68 (0,63 a 0,74)	0,73 (0,67-0,77)
Observador 6	0,73 (0,68 a 0,79)	0,77 (0,70 a 0,83)
Reporte de la mayoría	0,80 (0,76 a 0,85)	0,85 (0,80 a 0,90)

^a Acuerdo intraobservador para los seis radiólogos según las 4 categorías.

^b Acuerdo intraobservador para el resultado dicotomizado (mama densa/no densa).

el conocimiento y la experiencia del profesional, mientras que el método automatizado utiliza una estrategia cuantitativa para determinar la categoría ACR. En segundo lugar, los procesos de desarrollo y validación de una herramienta de *software* determinarán su desempeño en los distintos escenarios. Sería interesante mediante futuros estudios explorar los motivos de esta diferencia.

Si bien este trabajo fue realizado en una única institución, la misma es de referencia y recibe a diario derivaciones de todo el país. Un estudio multicéntrico sería conveniente para evaluar nuevas tecnologías. En relación con la muestra utilizada, esta cuenta con un bajo número de casos con patrón mamográfico extremadamente denso (ACR-d), acorde con la prevalencia reportada en el servicio en los últimos 5 años, durante los cuales los registros rondaron el 1-2%. Aun así, la prevalencia total de alta densidad (categorías c y d) en la muestra fue del 41%, también acorde con la prevalencia en nuestra población hospitalaria. El número de mamografías empleadas y la aleatorización del orden de los estudios evitó sesgos de memoria por posibles efectos de familiarización con la muestra de mamografías entre los dos tiempos de observación. Por último, nuestro diseño garantizó el ciego de los médicos a los reportes del *software* de clasificación automática y los diagnósticos del resto de los evaluadores.

Conclusión

Aunque existe una variabilidad general entre observadores e incluso para el mismo operador, la clasificación cualitativa de densidad mamaria es un método aceptable con un acuerdo entre moderado y casi perfecto inter- e intraobservador según los criterios establecidos en la 5.ª ed. del Atlas BI-RADS®. Hemos observado nivel de acuerdo moderado entre el reporte de los especialistas y un *software* disponible comercialmente. Futuros estudios podrán ahondar en la caracterización del acuerdo entre especialistas y métodos de clasificación automatizados.

Autoría

1. Responsable de la integridad del estudio: SB, KP.
2. Concepción del estudio: MT, KP, JF.

3. Diseño del estudio: BB, JF.
4. Obtención de los datos: MJC, MPS.
5. Análisis e interpretación de los datos: BB, JF, MT, KP.
6. Tratamiento estadístico: BB, MT.
7. Búsqueda bibliográfica: JF, MT, KP, MJC.
8. Redacción del trabajo: JF, MT, KP, BB, MPS.
9. Revisión crítica del manuscrito con aportaciones intelectualmente relevantes: SB.
10. Aprobación de la versión final: JF, MT, KP, BB, MPS, SB.

Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

Bibliografía

1. Winkler NS, Raza S, Mackesy M, Birdwell RL. Breast density: clinical implications and assessment methods. *Radiographics*. 2015;35:316–24. <http://dx.doi.org/10.1148/rg.352140134>.
2. Sickles EA, D'Orsi CJ, Bassett LW. ACR BI-RADS® Mammography. En: En: ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System. Reston, VA: American College of Radiology; 2013.
3. Ciatto S, Visioli C, Paci E, Zappa M. Breast density as a determinant of interval cancer at mammographic screening. *Br J Cancer*. 2004;90:393–6. <http://dx.doi.org/10.1038/sj.bjc.6601548>.
4. Wanders JOP, Holland K, Karssemeijer N, Peeters PHM, Veldhuis WB, Mann RM, et al. The effect of volumetric breast density on the risk of screen-detected and interval breast cancers: a cohort study. *Breast Cancer Res*. 2017;19:67. <http://dx.doi.org/10.1186/s13058-017-0859-9>.
5. Strand F, Azavedo E, Hellgren R, Humphreys K, Eriksson M, Shepherd J, et al. Localized mammographic density is associated with interval cancer and large breast cancer: a nested case-control study. *Breast Cancer Res*. 2019;21:8. <http://dx.doi.org/10.1186/s13058-019-1099-y>.
6. McCormack VA, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev*. 2006;15:1159–69. <http://dx.doi.org/10.1158/1055-9965.EPI-06-0034>.
7. Boyd NF, Guo H, Martin LJ, Sun L, Stone J, Fishell E, et al. Mammographic density and the risk and detection of breast cancer. *N Engl J Med*. 2007;356:227–36. <http://dx.doi.org/10.1056/NEJMoa062790>.

8. Swann CA, Kopans DB, McCarthy KA, White G, Hall DA. Mammographic density and physical assessment of the breast. *AJR Am J Roentgenol.* 1987;148:525–6, <http://dx.doi.org/10.2214/ajr.148.3.525>.
9. Mousa DSAL, Ryan EA, Mello-Thoms C, Brennan PC. What effect does mammographic breast density have on lesion detection in digital mammography? *Clin Radiol.* 2014;69:333–41, <http://dx.doi.org/10.1016/j.crad.2013.11.014>.
10. Carreira Gómez MC, Estrada Blan MC. What we need to know about dense breasts: implications for breast cancer screening. *Radiologia.* 2016;58:421–6, <http://dx.doi.org/10.1016/j.rx.2016.08.002>.
11. Sprague BL, Conant EF, Onega T, Garcia MP, Beaber EF, Herschorn SD, et al. Variation in Mammographic Breast Density Assessments Among Radiologists in Clinical Practice: A Multi-center Observational Study. *Ann Intern Med.* 2016;165:457–64, <http://dx.doi.org/10.7326/M15-2934>.
12. Eom HJ, Cha JH, Kang JW, Choi WJ, Kim HJ, Go E. Comparison of variability in breast density assessment by BI-RADS category according to the level of experience. *Acta Radiol.* 2018;59:527–32, <http://dx.doi.org/10.1177/0284185117725369>.
13. Alikhassi A, Esmaili Gourabi H, Baikpour M. Comparison of inter- and intra-observer variability of breast density assessments using the fourth and fifth editions of Breast Imaging Reporting and Data System. *Eur J Radiol Open.* 2018;5:67–72, <http://dx.doi.org/10.1016/j.ejro.2018.04.002>.
14. Melnikow J, Fenton JJ, Whitlock EP, Miglioretti DL, Weyrich MS, Thompson JH, et al. Supplemental Screening for Breast Cancer in Women With Dense Breasts: A Systematic Review for the U.S. Preventive Services Task Force. *Ann Intern Med.* 2016;164:268–78, <https://doi.org/10.7326/M15-1789>.
15. Jeffers AM, Sieh W, Lipson JA, Rothstein JH, McGuire V, Whittemore AS, et al. Breast Cancer Risk and Mammographic Density Assessed with Semiautomated and Fully Automated Methods and BI-RADS. *Radiology.* 2017;282:348–55, <http://dx.doi.org/10.1148/radiol.2016152062>.
16. Ciatto S, Bernardi D, Calabrese M, Durando M, Gentilini MA, Mariscotti G, et al. A first evaluation of breast radiological density assessment by QUANTRA software as compared to visual classification. *Breast.* 2012;21:503–6, <http://dx.doi.org/10.1016/j.breast.2012.01.005>.
17. Alonzo-Proulx O, Jong RA, Yaffe MJ. Volumetric breast density characteristics as determined from digital mammograms. *Phys Med Biol.* 2012;57:7443–57, <http://dx.doi.org/10.1088/0031-9155/57/22/7443>.
18. Martínez Gómez I, Casals El Busto M, Antón Guirao J, Ruiz Perales F, Llobet Azpitarte R. Semiautomatic estimation of breast density with DM-Scan software. *Radiologia.* 2014;56:429–34, <http://dx.doi.org/10.1016/j.rx.2012.11.007>.
19. Gao J, Warren R, Warren-Forward H, Forbes JF. Reproducibility of visual assessment on mammographic density. *Breast Cancer Res Treat.* 2008;108:121–7, <http://dx.doi.org/10.1007/s10549-007-9581-0>.
20. Luna D, Plazzotta F, Otero C, González Bernaldo de Quirós F, Baum A, Benítez S. Incorporación de tecnologías de la información y de las comunicaciones en el Hospital Italiano de Buenos Aires. Naciones Unidas: Santiago de Chile;. 2012. Disponible en: <http://hdl.handle.net/11362/3959>.
21. https://www.dma.no/files/298/fujifilm_amulet_innovality.pdf.
22. Ekpo EU, Ujong UP, Mello-Thoms C, McEntee MF. Assessment of Interradiologist Agreement Regarding Mammographic Breast Density Classification Using the Fifth Edition of the BI-RADS Atlas. *AJR Am J Roentgenol.* 2016;206:1119–23, <http://dx.doi.org/10.2214/AJR.15.15049>.
23. Cohen J. A Coefficient of Agreement for Nominal Scales. *Educ Psychol Meas.* 1960;20:37–46. Disponible en: <http://journals.sagepub.com/doi/10.1177/001316446002000104>.
24. Fleiss JL, Cohen J. The Equivalence of Weighted Kappa and the Intraclass Correlation Coefficient as Measures of Reliability. *Educ Psychol Meas.* 1973;33:613–9. Disponible en: <http://journals.sagepub.com/doi/10.1177/001316447303300309>.
25. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159–74, <https://www.ncbi.nlm.nih.gov/pubmed/843571>.
26. Berg WA, Zhang Z, Lehrer D, Jong RA, Pisano ED, Barr RG, et al. Detection of breast cancer with addition of annual screening ultrasound or a single screening MRI to mammography in women with elevated breast cancer risk. *JAMA.* 2012;307:1394–404, <http://dx.doi.org/10.1001/jama.2012.388>.
27. Berg WA, Campassi C, Langenberg P, Sexton MJ. Breast Imaging Reporting and Data System: inter- and intraobserver variability in feature analysis and final assessment. *AJR Am J Roentgenol.* 2000;174:1769–77, <http://dx.doi.org/10.2214/ajr.174.6.1741769>.
28. Gard CC, Aiello Bowles EJ, Miglioretti DL, Taplin SH, Rutter CM. Misclassification of Breast Imaging Reporting and Data System (BI-RADS) Mammographic Density and Implications for Breast Density Reporting Legislation. *Breast J.* 2015;21:481–9, <http://dx.doi.org/10.1111/tbj.12443>.
29. Ciatto S, Houssami N, Apruzzese A, Bassetti E, Brancato B, Carozzi F, et al. Categorizing breast mammographic density: intra- and interobserver reproducibility of BI-RADS density categories. *Breast.* 2005;14:269–75, <http://dx.doi.org/10.1016/j.breast.2004.12.004>.
30. Redondo A, Comas M, Macià F, Ferrer F, Murta-Nascimento C, Maristany MT, et al. Inter- and intraradiologist variability in the BI-RADS assessment and breast density categories for screening mammograms. *Br J Radiol.* 2012;85:1465–70, <http://dx.doi.org/10.1259/bjr/21256379>.
31. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159–74.
32. Tourassi GD, Floyd CE. The effect of data sampling on the performance evaluation of artificial neural networks in medical diagnosis. *Med Decis Making.* 1997;17:186–92, <http://dx.doi.org/10.1177/0272989X9701700209>.