



UNIVERSIDAD AUTÓNOMA DE SAN LUIS POTOSÍ

FACULTAD DE MEDICINA

HOSPITAL REGIONAL DE ALTA ESPECIALIDAD DR. IGNACIO MORONES PRIETO

Trabajo de investigación para obtener el diploma en la especialidad de Imagen y Radiología  
Diagnóstica y Terapéutica

**Evaluación de la variabilidad interobservador en la clasificación de calcificaciones  
mamarias mediante el sistema BI-RADS 5ta edición.**

**Dr. Abel de Jesús Guerrero Jaime**

DIRECTOR CLÍNICO

Dra. Rosario Margot Camargo Zebadúa

DIRECTOR METODOLÓGICO

Dr. Juan Manuel Shiguetomi Medina

Febrero 2025





UNIVERSIDAD AUTÓNOMA DE SAN LUIS POTOSÍ

FACULTAD DE MEDICINA

HOSPITAL REGIONAL DE ALTA ESPECIALIDAD DR. IGNACIO  
MORONES PRIETO

Trabajo de investigación para obtener el diploma en la especialidad de  
Imagen y Radiología Diagnóstica y Terapéutica

**Evaluación de la variabilidad interobservador en la clasificación de  
calcificaciones mamarias mediante el sistema BI-RADS 5ta edición.**

**Dr. Abel de Jesús Guerrero Jaime**

DIRECTOR CLÍNICO

Dra. Rosario Margot Camargo Zebadúa

DIRECTOR METODOLÓGICO

Dr. Juan Manuel Shiguetomi Medina

No. de CVU del CONACYT 313177

Identificador de ORCID 0000-0003-4131-093X

## SINODALES

Dr. Jorge Guillermo Reyes Vaca  
Médico Radiólogo.  
Neuro radiólogo.  
Presidente

---

Dr. Carlos Lambert Cerda  
Médico Radiólogo  
Sinodal

---

Dra. Zenyazen Cisneros Mejía  
Médico Radiólogo  
Sinodal

---

Dr. Ramsés Alejandro Padrón Hernández  
Médico Radiólogo  
Suplente

---



Evaluación de la variabilidad interobservador en la clasificación de calcificaciones mamarias mediante el sistema BI-RADS 5ta edición. © 2025 Por Abel de Jesús Guerrero Jaime. [Se distribuye bajo Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Febrero 2025

## **RESUMEN**

La variabilidad interobservador en la clasificación de calcificaciones mamarias puede afectar la precisión diagnóstica y la toma de decisiones clínicas. Este estudio tiene como objetivo evaluar la concordancia entre dos radiólogos con diferente nivel de experiencia en la clasificación de calcificaciones mediante el sistema BIRADS quinta edición utilizando el coeficiente kappa de Cohen.

Se realizó un estudio observacional, transversal y analítico en el Hospital Regional de Alta Especialidad Dr. Ignacio Morones Prieto. Se seleccionaron 100 mamografías con calcificaciones de pacientes mayores de 40 años, excluyendo aquellas con mala técnica o antecedentes de mastectomía. Dos radiólogos, uno con alta especialidad en imagen mamaria y otro residente en formación, evaluaron de manera independiente las imágenes. Se analizó la concordancia en tres categorías: calcificaciones benignas, sospechosas y categoría BIRADS.

Los resultados mostraron una concordancia moderada en la clasificación de calcificaciones benignas ( $\kappa=0.54$ ), lo que indica cierto grado de discrepancia entre observadores. La clasificación de calcificaciones sospechosas presentó una concordancia ligera ( $\kappa=0.20$ ) sugiriendo una mayor subjetividad en la evaluación de estos hallazgos. En contraste la clasificación BIRADS final alcanzó una concordancia casi perfecta ( $\kappa=0.87$ ) lo que demuestra que a pesar de las diferencias en la interpretación de hallazgos específicos los radiólogos logran un consenso en la categoría global del estudio.

Estos hallazgos resaltan la importancia de mejorar la reproducibilidad en la evaluación de calcificaciones mamarias especialmente en aquellas con características sospechosas. Se sugiere el desarrollo de estrategias de capacitación y el uso de herramientas de apoyo diagnóstico para optimizar la concordancia interobservador y mejorar la precisión en la detección del cáncer de mama.

## **PALABRAS CLAVE**

Variabilidad interobservador, calcificaciones mamarias, Breast Imaging Reporting and Data System (BIRADS).

## ÍNDICE

	Página
Resumen .....	5
Índice .....	6
Lista de abreviaturas .....	8
Lista de definiciones .....	9
Dedicatorias .....	10
Antecedentes .....	11
Justificación .....	15
Hipótesis .....	16
Objetivos .....	16
Sujetos y métodos .....	16
Análisis estadístico .....	18
Ética .....	18
Resultados .....	19
Discusión .....	20

Limitaciones y/o nuevas perspectivas de investigación .....	21
Conclusiones .....	21
Bibliografía .....	22
Anexo 1 ( Valoración del coeficiente kappa ) .....	24

### LISTA DE CUADROS

	Página
Cuadro 1. (Resultados de coeficiente kappa calculado y nivel de concordancia por rubro.) .....	19

## LISTA DE ABREVIACIONES Y SÍMBOLOS

- **BI-RADS:** Breast Imaging Reporting and Data System
- **CDIS:** Carcinoma ductal in situ
- **FFDM:** Mamografía digital de campo completo
- **2D-SM:** Mamografía sintetizada en 2D
- **DBST:** Tomografía computarizada dedicada a la mama
- **DBT:** Tomosíntesis digital mamaria
- **CAD:** Detección asistida por computadora
- **PPV:** Valor predictivo positivo
- **DR:** Radiografía digital
- **CR:** radiografía computarizada



## LISTA DE DEFINICIONES

- **Mamografía:** Examen radiológico de la mama.
- **Variabilidad interobservador:** La incapacidad del observador para medir o identificar un fenómeno con precisión, lo que da lugar a un error.
- **Calcificación fisiológica:** Proceso por el cual el tejido orgánico se endurece por el depósito fisiológico de sales de calcio.
- **Calcificación patológica:** Depósito patológico de sales de calcio en los tejidos.

## **DEDICATORIAS**

A mi familia por acompañarme en este largo camino de formación médica.

A mis maestros por brindarme sus conocimientos y experiencias.

A mis compañeros por ser parte fundamental en estos 3 años de residencia.

A Lilí por ser mi motor durante todos estos años y tenerme la paciencia que la residencia conlleva.

A Archie por ser el mejor acompañante que existe.

## ANTECEDENTES

La clasificación de calcificaciones en mastografías es un aspecto fundamental en la detección temprana del cáncer de mama. La detección de las calcificaciones es importante, ya que aunque, llegan a ser benignas, en algunos casos pueden indicar la presencia de carcinoma ductal in situ (CDIS) o cáncer invasivo. En la interpretación de estas imágenes se ha implementado el sistema BI-RADS (Breast Imaging Reporting and Data System) para estandarizar la clasificación de las lesiones y así poder reducir la variabilidad interobservador, sin embargo, a pesar de estas herramientas la variabilidad interobservador sigue siendo un desafío en la actualidad, especialmente cuando se trata de clasificar calcificaciones (1).

En un estudio que comparó la mamografía digital de campo completo (FFMD) con la mamografía sintetizada en 2D (2D-SM), se encontró que las dos modalidades tienen capacidad similar en la detección de calcificaciones, la calidad de la imagen de la FFMD es ligeramente mas superior a la 2D-SM, lo que podría influir en la interpretación entre los diferentes observadores. Además la exposición a la radiación fue menor en la mamografía sintetizada, lo que pudiera ser una alternativa en ciertos contextos clínicos (1).

El uso del BI-RADS no se limita únicamente a las mamografías tradicionales, si no que también ha demostrado ser útil en otras modalidades de imagen como en la tomografía computarizada dedicada a la mama (DBCT). Un estudio evaluó el impacto de utilizar un atlas BI-RADS en la clasificación de lesiones detectadas con DBCT, aunque la evaluación de calcificaciones mediante esta técnica sigue representando un desafío debido a la baja resolución de este método (2).

Otro estudio evaluó la relación entre las características histopatológicas del cancer de mama y la rigidez de los tumores mediante el uso de elastografía por esfuerzo. Se encontró que la rigidez tumoral muestra variación en su rigidez según el tipo histológico del cáncer, pero hubo una diferencia significativa en la elasticidad de los tumores con o sin calcificaciones, esto sugiere que aunque las calcificaciones sin un factor importante en la detección de los tumores, otros datos como el grado histológico pueden jugar un papel más importante en la evaluación de la malignidad de los tumores, factor que podría explicar la variabilidad en las evaluaciones entre radiólogos (3).

La tomosíntesis digital mamaria (DBT) ha sido una de las tecnologías mas prometedoras para mejora la detección de lesiones mamarias, particularmente en mujeres con mamas densas. Un estudio comparó el rendimiento diagnóstico de la DBT con un ángulo de escaneo amplio frente a la mamografía digital de campo completo (FFDM) y encontró que la DBT ofrece una mejor sensibilidad para la detección de lesiones malignas, sin embargo a pesar de esta mejora tecnológica la

especificidad fue ligeramente inferior, lo que sugiere que si bien la DBT puede reducir la variabilidad en la detección de lesiones malignas aun existen desafíos en la interpretación precisa de lesiones sospechosas y particularmente en las calcificaciones (4).

La variabilidad entre los radiólogos en la interpretación de imágenes mamográficas ha sido estudiada ampliamente, un estudio investigó la variabilidad interobservador en el uso de los descriptores BI-RADS mostrando que la concordancia es alta para identificar masas y calcificaciones, pero baja para asimetrías en una sola vista y distorsiones de la arquitectura. Este tipo de hallazgos son especialmente relevantes en el contexto de las microcalcificaciones ya que pequeñas diferencias en la interpretación pueden condicionar decisiones clínicas muy diferentes, como la recomendación de biopsias y vigilancia activa (5).

Un aspecto importante que afecta la variabilidad interobservador es la experiencia del radiólogo, un estudio reveló que los médico radiólogos con más de 2500 mamografías leídas durante su carrera tienen mayor concordancia en la clasificación final de las mamografías que aquellos con menos experiencia, esto sugiere que la cantidad de casos interpretados es un factor clave para mejora la precisión en la clasificación de calcificaciones y otras lesiones, lo que puede ser un área de mejora a través de programas de formación intensiva para los radiólogos (6).

La introducción de segundas opiniones ha sido una estrategia propuesta para intentar mejorar la presión en el diagnóstico en las mamografías. Un estudio examinó el impacto de la revisión de una junta multidisciplinaria en la clasificación BI-RADS de lesiones no palpable, lo que mostró que una segunda opinión mejoraba la precisión en el diagnóstico, aumentando la tasa de detección de lesiones malignas y de alto riesgo. Esto sugiere que la revisión por un grupo de expertos puede ser una estrategia eficaz para reducir la variabilidad interobservador (7).

El uso de arbitraje por un tercer lector en lecturas dobles discrepantes ha sido otro enfoque que se ha investigado para reducir la variabilidad interobservador. Un estudio analizó el arbitraje en mamografías digitales de tamizaje y mostró que, aunque esta técnica reducía la tasa de derivación y mejoraba el valor predictivo positivo (PPV), también disminuía ligeramente la sensibilidad del programa de cribado, esto plantea un dilema en como equilibrar la reducción de falsos positivos con la necesidad de mantener una alta tasa de sensibilidad en la detección del cáncer (8).

La variabilidad interobservador también se ha evaluado utilizando métricas estadísticas como lo es el coeficiente *Kappa*, que corrige el acuerdo por azar. Un estudio detallado sobre la medición del acuerdo entre observadores encontró que

aunque un alto nivel de acuerdo entre observadores no garantiza una alta precisión, las medidas de concordancia siguen siendo útiles para evaluar la fiabilidad de los métodos diagnósticos (9).

La confiabilidad de las primeras impresiones de los radiólogos también ha sido objeto de estudio. Se evaluó la concordancia de las primeras impresiones de la clasificación de microcalcificaciones de sospecha revelando una variabilidad moderada en los resultados con un índice Kappa de 0.45, esto indica que aunque los radiólogos tienden a estar de acuerdo en casos de lesiones claramente benignas o malignas, la variabilidad aumenta en aquellos casos en los que las calcificaciones que presentan características intermedias lo que puede dar lugar a discrepancias en la recomendación de biopsias o seguimientos. Este estudio sugiere que el uso de herramientas adicionales así como la detección asistida por computadora (CAD) o las segundas opiniones podrían ayudar a mejorar la precisión diagnóstica (10).

Un estudio sobre la variabilidad en los resultados de tamizaje entre pares de radiólogos, se observa que la tasa de detección de cáncer y la tasa de derivación variaban significativamente entre las diferentes combinaciones de radiólogos. Las tasas de detección oscilaban entre 4.0 y 6.3 por cada 1000 mamografías, mientras que las tasas de derivación variaban entre 1.0% y 1.5%, estas diferencias sugieren que las combinaciones de radiólogos con diferentes niveles de experiencia pueden influir en los resultados de cribado (11).

La variabilidad en la detección del cáncer de mama está influenciada por el tipo de tecnología de detección utilizada. Un estudio comparó el rendimiento de cuatro tipos diferentes de detectores de mamografía que incluyeron la radiología digital (DR) con selenio y cesio, y dos tipos de radiografías computarizadas (CR) con fosfato y en aguja. Los resultados demostraron que los detectores DR ofrecieron tasas más altas de detección de cáncer en comparación con los detectores CR. Además los detectores CR mostraron ser significativamente menos precisos para la identificación de clusters de microcalcificaciones malignas con una disminución del 22% en su tasa de detección comparado con los detectores DR, esto resalta la necesidad de utilizar tecnología de alta calidad, ya que la diferencia de rendimiento de los detectores puede ser un factor determinante en la precisión del diagnóstico (12).

La vista de magnificación geométrica ha demostrado ser una herramienta valiosa para mejorar la caracterización de las microcalcificaciones en mamografías digitales. Un estudio evaluó el impacto de la vista de magnificación en la interpretación de microcalcificaciones sospechosas y encontró que la adición de estas vistas mejoró significativamente la visibilidad de las mismas en el 76% de los casos. Esto permitió una mejor diferenciación entre calcificaciones benignas y

malignas y esto influyó en decisiones clínicas como la recomendación de biopsias y seguimientos (13).

A pesar de los avances tecnológicos, un problema recurrente es la omisión de las microcalcificaciones en las mamografías. Un estudio retrospectivo analizó la causa de la omisión de microcalcificaciones durante mastografías de tamizaje encontrando que el error perceptivo fue la principal causa. El 45% de los casos de carcinoma ductal in situ (CDSI) con microcalcificaciones fueron pasados por alto por menos uno de los lectores y la mayoría de estos errores fueron perceptivos en lugar de interpretativos (14).

En cuanto a la confiabilidad del sistema BI-RADS en la evaluación final de las mamografías, un estudio sobre la variabilidad interobservador demostró que aunque el sistema es efectivo para estandarizar la clasificación de lesiones mamarias existe una variabilidad considerable entre observadores en la interpretación de casos más ambiguos como las categorías BI-RADS 3 y 4. La concordancia interobservador fue moderada ( $\kappa = 0.48$ ), lo que indica que aunque el sistema BI-RADS ha mejorado la consistencia en la interpretación de las imágenes la variabilidad continua siendo un desafío (15).

En conclusión, la variabilidad en la interpretación de las mamografías representa una problemática importante a pesar de los avances tecnológicos y la implementación de sistemas de estandarización como el BI-RADS. La elección del tipo de detector, el uso de vistas de magnificación y la corrección de errores perceptivos son factores clave que pueden mejorar la precisión en la detección de calcificaciones y microcalcificaciones. La presente tesis busca abordar estos problemas proporcionando un análisis detallado de la variabilidad interobservador en la clasificación de calcificaciones mamarias de acuerdo con el sistema BI-RADS y contribuir al desarrollo de estrategias que mejoren la precisión diagnóstica en la detección del cáncer de mama.

## **JUSTIFICACIÓN**

La variabilidad interobservador en la interpretación de las calcificaciones mamarias tiene un impacto significativo en el manejo clínico de las pacientes. Esta variabilidad puede llevar a diagnósticos incorrectos y tratamiento inadecuados, lo que recalca la necesidad de realizar investigaciones que aborden este crítico problema. La presente tesis se justifica en la necesidad de evaluar la variabilidad interobservador en la clasificación de calcificaciones mamarias mediante el sistema BI-RADS 5ta edición y explorar estrategias para reducir esta variabilidad optimizando así el diagnóstico y manejo de las pacientes.

Este estudio contribuirá a la literatura existente a proporcionar una evaluación integral de la variabilidad interobservador en un contexto clínico específico, utilizando tecnologías avanzadas y enfoques colaborativos. Además los hallazgos que se obtengan pueden informar el desarrollo de guías y programas de formación así como la implementación de nuevas tecnologías para mejorar la precisión diagnóstica y manejo del cáncer de mama.

## **PREGUNTA DE INVESTIGACIÓN**

¿Cuál es la variabilidad interobservador en la clasificación de calcificaciones mamarias mediante el sistema BI-RADS 5ta edición en mamografías?

## **HIPÓTESIS**

Existe una variabilidad significativa entre radiólogos de diferentes niveles de experiencia en la clasificación de calcificaciones mamarias mediante el sistema BI-RADS 5ta edición.

## **OBJETIVOS**

- Objetivo general
  - Evaluar la variabilidad interobservador en la clasificación de calcificaciones mamarias mediante el sistema BI-RADS 5ta edición.
  
- Objetivos específicos
  - Cuantificar la concordancia entre radiólogos de diferentes niveles de experiencia en la clasificación de calcificaciones.
  - Identificar los factores que contribuyen a la variabilidad interobservador en la interpretación de calcificaciones mamarias.

## **SUJETOS Y MÉTODOS**

**Tipo de estudio:** Observacional, transversal y analítico.

**Metodología:** El estudio se llevó a cabo en el Hospital Regional de Alta Especialidad Dr. Ignacio Morones Prieto, donde se cuenta con un amplio acceso a mastografías y un equipo multidisciplinario de radiólogos con diversos niveles de experiencia.

Se seleccionaron mamografías de pacientes que se realizaron durante los años 2023-2024 con calcificaciones mamarias. Se incluyeron tanto pacientes con antecedentes de cáncer de mama como aquellas sin antecedentes conocidos.

### **Criterios de selección:**

- Criterios de inclusión: Mamografías que presenten calcificaciones realizadas a pacientes femeninos mayores de 40 años.
- Criterios de exclusión: Mamografías de mala calidad técnica que impidan una adecuada interpretación y pacientes con mastectomías previas.



## VARIABLES EN EL ESTUDIO:

<b>Dependientes</b>				
<b>Variable</b>	<b>Definición operacional</b>	<b>Valores posibles</b>	<b>Unidades</b>	<b>Tipo de variable</b>
Clasificación de las calcificaciones mamarias (según BIRADS 5ta edición)	Resultado de la clasificación de las calcificaciones mamarias en la mastografía de acuerdo con los criterios establecidos por BIRADS 5ta edición	-BIRADS 1. Negativa -BIRADS 2: Benigna -BIRADS 3: Probablemente benigna -BIRADS 4: Sospechosa de malignidad -BIRADS 5: Altamente sospechosa de malignidad	Nominal/Categórica	Ordinal
<b>Independientes</b>				
Experiencia del radiólogo	Nivel de experiencia del radiólogo que realiza la interpretación de las calcificaciones mamarias determinado por los años de práctica o nivel de formación	-Principiante: menos de 5 años de experiencia -Intermedio: 5 a 10 años de experiencia -Avanzado: más de 10 años de experiencia	Años de experiencia/Categorías por nivel	Ordinal
Densidad mamaria	Clasificación de la densidad del tejido mamario en la mastografía según el sistema BIRADS	-A: Mayormente grasa -B: Fibroglandular disperso -C: Heterogéneamente denso -D: Extremadamente denso	Nominal/Categórica	Ordinal
Características de las calcificaciones	Descripción de las propiedades de las calcificaciones mamarias que incluyen tamaño, forma y distribución	-Tamaño: Microcalcificaciones (<0.5 mm) o Macrocalcificaciones (>0.5 mm). -Forma: Redondas, irregulares, en racimo, lineales. -Distribución: Agrupadas, segmentarias, regionales	Nominal/Categórica	Ordinal
<b>VARIABLES DE CONTROL (CONFUSORAS)</b>				
Calidad de la imagen	Evaluación subjetiva u objetiva de la calidad técnica de la imagen obtenida en la mastografía	-Baja calidad -Calidad aceptable -Alta calidad	Categórica	Ordinal

- Variables dependientes: Clasificación de las calcificaciones mamarias según el sistema BIRADS 5ta edición.
- Variables independientes: Experiencia del radiólogo, densidad mamaria, calidad de la imagen y características de las calcificaciones.

### **Tipo de muestreo**

Se empleó un muestre aleatorio simple para seleccionar las mamografías de pacientes que cumplan los criterios de inclusión y exclusión establecidos.

### **Cálculos del tamaño de la muestra**

En 1986 Fleiss, Cohen y Everitt definieron los niveles de concordancia de diagnóstico para estudios de radiología, por lo que el tamaño de la muestra fue calculado de acuerdo a lo sugerido, donde para estudios de concordancia entre dos radiólogos una muestra de al menos 100 estudios es suficiente para calcular índice Kappa confiable (16).

### **ANÁLISIS ESTADÍSTICO**

Se considerará la evaluación independiente de dos radiólogos con diferente nivel de experiencia, uno médico radiólogo con alta especialidad en imagen e intervención de la mama y otro cursando el último de la residencia de la especialidad de imagenología diagnóstica y terapéutica. Para cuantificar el grado de acuerdo entre ambos observadores se empleará el coeficiente kappa de Cohen.

La interpretación del coeficiente kappa se basará en la escala de Landis y Koch (1977), que categoriza la concordancia de la siguiente manera: <0.00 (concordancia menor al azar), 0.00-0.20 (ligera), 0.21-0.40 (aceptable), 0.41-0.60 (moderada), 0.61-0.80 (considerable) y 0.81-1.00 (casi perfecta). ANEXO 1.

### **ÉTICA**

El presente estudio no implica una intervención directa en los sujetos lo que reduce de manera significativa el riesgo físico o psicológico, sin embargo sigue siendo fundamental garantizar la confidencialidad de los datos. Toda la información utilizada se encuentra debidamente codificada para evitar la identificación de los pacientes respetando así su privacidad.

El estudio se considera una investigación sin riesgo que no requiere obtención de consentimiento informado. Se realiza en cumplimiento de los establecido en el Reglamento de la Ley General de Salud en Materia de Investigación. Todos los procedimientos se apegan a la NOM-012-SSA-2012 que establece los criterios para la ejecución de proyectos de investigación para la salud en seres humanos, así como a la declaración de Helsinki y sus enmiendas.

## RESULTADOS

Tras haber finalizado la revisión de los casos y clasificación de manera individual entre los dos observadores se realizó el análisis en tres rubros, clasificación de calcificaciones benignas, clasificación de calcificaciones sospechosas y clasificación en la categoría BIRADS final, con lo que se obtuvieron los siguientes resultados.

### Clasificación de calcificaciones benignas

Se evaluó el grado de acuerdo entre los dos radiólogos en la clasificación de calcificaciones dentro del grupo de las benignas

Se calculó un coeficiente kappa de 0.54 lo que indica una concordancia moderada entre los observadores.

### Clasificación de calcificaciones sospechosas

En esta fase se analizó la concordancia en las calcificaciones dentro del grupo de las sospechosas calculándose un coeficiente kappa de 0.20 lo que indica una concordancia ligera .

### Clasificación de la categoría BIRADS

Finalmente se evaluó al concordancia en la clasificación BIRADS del estudio obteniendo un coeficiente kappa de 0.87, lo que representa una concordancia casi perfecta entre los dos observadores.

**Cuadro 1.** Resultados de coeficiente kappa calculado y nivel de concordancia por rubro.

Rubro	Coeficiente kappa	Nivel de concordancia
Calcificaciones benignas	0.54	Moderada
Calcificaciones sospechosas	0.20	Ligera
Categoría BIRADS	0.87	Casi perfecta

## DISCUSIÓN

Los resultados obtenidos reflejan diferencias en la interpretación de las calcificaciones benignas y sospechosas.

La concordancia en la clasificación de calcificaciones benignas se encontró en un nivel moderado ( $\kappa = 0.54$ ), lo que sugiere que aunque exista un grado aceptable de acuerdo todavía se observan discrepancias en la identificación de estos patrones. Este hallazgo es consistente con estudios previos que han reportado cierta variabilidad en la clasificación de calcificaciones benignas. La experiencia del radiólogo puede influir en la interpretación de estos hallazgos generando diferencias en la categorización.

Por otro lado, la clasificación de calcificaciones sospechosas mostró menor concordancia entre los observadores, con un coeficiente kappa de 0.20 que indica una concordancia ligera. Este resultado cobra importante relevancia ya que estos resultados sugieren que la evaluación de las calcificaciones con características de sospecha o atípicas presenta una alta subjetividad y depende en gran medida del criterio individual del radiólogo, además de que la menor experiencia en la evaluación de estas lesiones podría llevar a mayor variabilidad en su interpretación y que impactaría en la toma de decisiones clínicas como en la necesidad de realizar estudios adicionales o biopsias.

La variabilidad en la clasificación de las calcificaciones sospechosas también puede atribuirse a la diversidad de patrones morfológicos y distribucionales descritos en el sistema BIRADS 5ta edición.

A pesar de la variabilidad observada en la evaluación de calcificaciones individuales, la concordancia en la clasificación del BIRADS final fue alta ( $\kappa 0.87$ ), lo que indica una concordancia casi perfecta entre los observadores. Este hallazgo es clave ya que demuestra que a nivel global los radiólogos logran llegar a un consenso en la categorización del estudio mamográfico lo cual es fundamental en la toma de decisiones clínicas.

La alta concordancia en la clasificación BIRADS sugiere que aunque puedan existir diferencias en la interpretación de hallazgos específicos los radiólogos tienden a utilizar criterios más generales relevantes para definir la categoría final del estudio. Esto se podría considerar como un aspecto positivo ya que a pesar de las discrepancias en la clasificación individual de las calcificaciones, la recomendación final para el manejo del paciente es relativamente consistente entre los observadores.

## **LIMITANTES Y NUEVAS PERSPECTIVAS DE INVESTIGACIÓN**

Es importante considerar que este estudio presenta algunas limitaciones. En primer lugar el tamaño muestral de 100 mastografías, aunque es representativo podría no reflejar completamente la variabilidad que se observa en la práctica clínica diaria. Además el estudio solo incluyó dos radiólogos con diferente nivel de experiencia, por lo que los resultados podrían diferir si se incluyera una muestra más amplia de especialistas con distintos niveles de formación.

Otra limitación es la ausencia de correlación con datos histopatológicos, si bien el objetivo del estudio fue evaluar la concordancia interobservador, no se analizó la relación entre la clasificación de los radiólogos y los resultados definitivos de biopsia, lo que podría aportar información adicional sobre la precisión diagnóstica de las evaluaciones realizadas.

## **CONCLUSIONES**

Este estudio evidencia que la concordancia interobservador en la clasificación de calcificaciones mamarias varían según el tipo de hallazgo evaluado. Mientras que la clasificación de calcificaciones típicamente benignas mostró una concordancia moderada y las sospechosas una concordancia baja, la determinación del BIRADS final tuvo un alto nivel de acuerdo, lo que sugiere que a nivel general los radiólogos logran llegar a una decisión diagnóstica coherente y reproducible.

Los resultados resaltan la importancia de implementar estrategias de formación continua y herramientas de apoyo diagnóstico para mejorar la reproducibilidad en la interpretación de calcificaciones mamarias. Optimizar la concordancia interobservador es clave para garantizar diagnósticos más precisos y mejorar la toma de decisiones clínicas en la detección temprana del cáncer de mama.

## BIBLIOGRAFÍA

1. Wahab RA, Lee SJ, Zhang B, Sobel L, Mahoney MC. A comparison of full-field digital mammograms versus 2D synthesized mammograms for detection of microcalcifications on screening. *European Journal of Radiology*. 2018 Oct; 107: 14-9.
2. Jung HK, Kuzmiak CM, Kim KW, Choi NM, Kim HJ, Langman EL, et al. Potential Use of American College of Radiology BI-RADS Mammography Atlas for Reporting and Assessing Lesions Detected on Dedicated Breast CT Imaging. *Academic Radiology*. 2017 nov;24(11):1395-401.
3. Durhan G, Öztekin PS, Ünverdi H, Degirmenci T, Durhan A, Karakaya J, et al. Do Histopathological Features and Microcalcification Affect the Elasticity of Breast Cancer? *Journal of Ultrasound in Medicine*. 2017 Feb 27;36(6):1101-8.
4. Clauser P, Nagl G, Helbich TH, Pinker-Domenig K, Weber M, Kapetas P, et al. Diagnostic performance of digital breast tomosynthesis with a wide scan angle compared to full-field digital mammography for the detection and characterization of microcalcifications. *European Journal of Radiology*. 2016 Dec;85(12):2161-8.
5. Lee AY, Winser DJ, Aminololama-Shakeri S, Arasu VA, Feig SA, Hargreaves J, et al. Inter-reader Variability in the Use of BI-RADS Descriptors for Suspicious Findings on Diagnostic Mammography *Academic Radiology*. 2017 Jan;24(1):60-6.
6. Abdelrahman MA, Rawashdeh MA, McEntee M, Abu Tahoun L, Brennan P. Career caseload predicts interobserver agreement on the final classification of a mammogram. *Journal of Medical Imaging and Radiation Oncology*. 2019 Feb;63(2):197-202.
7. De Margerie-Mellon C, Debry JB, Dupont A, Cuvier C, Giacchetti S, ,Teixeira L, et al. Nonpalpable breast lesion: impact of a second-opinion review at a breast unit on BI-RADS classification. *European Radiology*. 2021 Jan 18;31(8):5913-23.
8. Klompenhouwer EG, Voogd AC, Gerard, Luc, Tjan-Heijnen VC, Mireille, et al. Discrepant screening mammography assessment at blinded and non-blinded double Reading: impact of a arbitration by a third reader on screening outcome. *European Radiology*. 2015 Apr 18;25(10):2821-9.

9. Kundel HL, Polansky M. Measurements of Observer Agreement. *Radiology*. 2003 Aug;228(2):303-8.
10. Ziba Gandomkar, Somphone Siviengphanom, Suleiman ME, Wong D, Reed W, Ekpo EU, et al. Reliability of radiologists' first impression when interpreting a screening mammogram. *PLOS ONE*. 2023 Apr 25;18(4):e0284605-5.
11. Klompenhouwer EG, Duijm LEM, Voogd AC, den Heeten GJ, Nederend J, Jansen FH, et al. Variations in screening outcome among pairs of screening radiologist at non-blinded double Reading of a screening mammograms: a population-based study. *European Radiology*. 2014 Feb 6;24(5):1097-104.
12. Mackenzie A, Warren LM, Wallis MG, Cooke J, Given-Wilson RM, Dance DR, et al. Breast cancer detection rates using four different types of mammography detectors. *European Radiology*. 2015 Jun 25;26(3):874-83.
13. Fallenberg E, Dimitrijevic L, Diekmann F, Diekmann S, Kettritz U, Poellinger A, et al. Impact of Magnification Views on the Characterization of Microcalcifications in Digital Mammography. *Röfo – Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren*. 2013 Sep 2;186(03):274-80.
14. Peters G, Jones CM, Daniels K. Why is microcalcification missed on mammography? *Journal of Medical Imaging and Radiation Oncology*. 2012 Dec 28;57(1):32-7.
15. Salazar AJ, Romero JA, Bernal OA, Moreno AP, Velasco SC, Reliability of the BIRADS Final Assessment Categories and Management Recommendations in a Telemammography Context. *Journal of the American College of Radiology*. 2017 May 1;14(5):668-692.
16. Fleiss JL, Cohen J, Everitt BS. Large sample standard errors of kappa and weighted kappa. *Psychol Bull* 1969;72:323-327.

## ANEXO 1

### Valoración del coeficiente kappa (Landis y Koch, 1977)

<b>Coeficiente kappa</b>	<b>Fuerza de la concordancia</b>
0,00	Pobre ( <i>Poor</i> )
0,01 - 0,20	Leve ( <i>Slight</i> )
0,21 - 0,40	Aceptable ( <i>Fair</i> )
0,41 - 0,60	Moderada ( <i>Moderate</i> )
0,61 - 0,80	Considerable ( <i>Substantial</i> )
0,81 - 1,00	Casi perfecta ( <i>Almost perfect</i> )