



**FACULTAD DE
MEDICINA**



**UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY**

Nuevos métodos para la evaluación de la calidad de imágenes de fantomas mamográficos de la ACR basados en la visión humana I: Exploración, descripción e hipótesis sobre los errores

AUTORES

Franco Hernandez, Gonzalo Martinez, Jessica Menezes, Susy Figueron

ORIENTADORES

Silvana Silveira, Horacio Botti

Tabla de contenidos:

1.	Resumen	5
2.	Introducción	6
3.	Marco teórico	
3.1.	Estadificación del cáncer de mama.....	8
3.2.	Avances del conocimiento y la atención clínica de mujeres con cáncer de mama.....	9
3.3.	Teorías y paradigmas biomédicos sobre el cáncer de mama.....	10
3.4.	Teorías de visión humana	12
3.5.	Teoría sobre la formación de imágenes de RX (sistemas lineales).....	13
3.6.	Paradigmas de evaluación de la exactitud de los juicios de los observadores sobre la detección y posición de microcalcificaciones en fantasmas mamográficos ACR.....	16
3.7.	Teorías sobre la detección de microcalcificaciones en imágenes 2D mamográficas.....	18
3.8.	Teorías sobre la información y la pérdida de exactitud.....	20
4.	Antecedentes	20
5.	Hipótesis, objetivos, preguntas de investigación	25
6.	Diseño	26
7.	Materiales y métodos	27
8.	Resultados	
8.1.	Evaluación de la posible utilidad del método 1.	29
8.2.	Características del proceso de registro realizado y del conjunto de datos obtenidos.....	33
8.3.	Análisis de datos: medidas de centralidad y de variabilidad de la posición.....	33
8.4.	Evaluación de la distribución de la variabilidad posicional y separación de errores.....	35
8.5.	Modelo que representa la hipótesis ajustada.....	36
8.6.	Variación intergrupar y calidad del ajuste del modelo estadístico.....	37
8.7.	Interpretación de los parámetros del modelo lineal horizontal.....	38
8.8.	Estimación del valor crítico (SD) e identificación de errores de registro.....	39
8.9.	Cálculo de valores críticos en Log10 y en pixeles del desvío estándar por grupo y eje.....	39
8.10.	Error posicional en X vs en Y.....	40

9.	Discusión.....	42
10.	Conclusión.....	45
11.	Conflictos de intereses.....	45
12.	Referencias bibliográficas.....	46
13.	Anexo.....	50

Resumen

Introducción: El tamizaje y el diagnóstico del cáncer de mama (CM) se benefician del mejor rendimiento clínico de los estudios mamográficos. Las mamografías (MX) de alta calidad pueden detectar detalles como microcalcificaciones pequeñas, las cuales pueden ser marcadores radiológicos de algunas formas de la enfermedad. Los fantasmas del Colegio Americano de Radiología (ACR) son ampliamente utilizados en el control de calidad de equipamientos para la acreditación de servicios de imagenología mamaria.

Antecedentes: Existen al menos dos métodos previos que se consideraron relevantes: el método de acreditación del ACR(1) y el método de evaluación de la calidad de imágenes de Huda(2).

Objetivos: Esta investigación busca proponer y evaluar preliminarmente nuevos métodos de análisis basados en la visión humana para el estudio de la calidad de imágenes de fantasmas mamográficos ACR. **Materiales y Métodos:** se idearon 4 métodos nuevos que utilizan información aportada por expertos humanos: ("N° Visible De Micros ESPERADA Por Grupo", " N° Visible De Micros ESPERADA Por Grupo e Ítem", " N° Visible De Micros Por Coordenadas", y "Método de error posicional"). Se utilizó una muestra de 155 imágenes extraídas de un juego original de la investigación de Maestría de la Lic. Maria Sarasua. Estas imágenes fueron obtenidas usando fantasmas ACR(Modelo 015), en tres servicios de mamografía del área metropolitana de Montevideo durante 2021 y 2022, variando

ampliamente en forma experimental los valores de los parámetros de adquisición. Cinco de los autores (cuatro estudiantes avanzados y un licenciado en Imagenología) visualizaron independientemente las imágenes con el objetivo de detectar y registrar las coordenadas en píxeles del punto de máxima intensidad para cada una de las microcalcificaciones simuladas en el fantoma. **Resultados:** Se produjo una matriz de $6 \times 5 \times 5 \times 155 = 23250$ registros (6 micros por cada uno de los 5 grupos a evaluar por 5 observadores en 155 imágenes). Estos datos dieron lugar a un conjunto de descripciones e hipótesis, algunas de las cuales se validaron internamente y otras con la literatura. Se valoró la presencia de errores de registro y se propuso un método de depuración ajustando un modelo mixto compuesto por una distribución normal más una recta horizontal a la distribución observada de frecuencias. Los resultados obtenidos mostraron una adecuada correspondencia entre el modelo teórico y los datos empíricos, de modo que se concluye que la mayor parte de la variabilidad se puede modelar como errores posicionales y una fracción menor, como errores de registro. Como se esperaba, los grupos con microcalcificaciones de menor tamaño presentan mayor error posicional, mientras que los grupos con señales más intensas alcanzan una mayor exactitud, lo que varió sistemáticamente entre los observadores, lo que sugiere fuertemente una la relación entre las evaluaciones de la

calidad de las imágenes y limitaciones perceptuales del observador. **Conclusiones:** Los métodos desarrollados integran principios psicofísicos y estadísticos en la evaluación de la calidad de imágenes mamográficas, siendo más exacta la evaluación que utiliza la información posicional cuando ella está disponible frente a los otros métodos conocidos. La investigación también aporta un marco teórico y conceptual para futuras investigaciones orientadas al mejor control de la calidad a partir de la percepción humana para la optimización de los sistemas de mamografía.

Introducción

El cáncer de mama destaca como la primera causa de muerte por cáncer en mujeres a nivel local, regional y mundial(3). Las estadísticas indican que en promedio 1 de cada 11 mujeres uruguayas desarrolla cáncer de mama en algún momento de su vida, siendo un importante problema de salud pública(4). Esta patología es originada por factores genéticos, ambientales, de estilo de vida y médicos, mostrando un origen en general multifactorial. Según el peso de la predisposición genética su origen etiológico se puede clasificar en cáncer de mama esporádico (no hereditario o no asociado a una predisposición genética heredada importante), 70-80 % de los casos; y familiar o hereditario, que se divide en dos categorías, con la alteración de diversos genes (casos poligénicos y oligogénicos) responsable del 15-20% y monogénico, que deriva de la mutación de un único gen, presentando entre un 5-10% de los casos. (3)

El 99% de los casos de cáncer de mama se presentan en mujeres(4). Un factor de riesgo importante son los antecedentes personales y/o familiares, predisposición genética los principales genes con predisposición para desarrollar estas patología son el BRCA1 y BRCA2(4). También existen fuertes evidencias de que la exposición a hormonas sexuales femeninas, progestágenos y estrógenos, juega un rol fundamental en el desarrollo de la enfermedad(5). El aumento de la esperanza de vida ha incrementado la prevalencia de enfermedades crónicas como el cáncer(6). El envejecimiento poblacional de Uruguay y de muchos otros países en todo el mundo también se debe a una disminución en la tasa de natalidad, resultando en una fase de crecimiento muy lento o estancado(7). El riesgo aumenta a edades más avanzadas, en nuestro país el 77% de los casos se presentan en mujeres de 50 años o más. Para disminuir la incidencia lo fundamental es promover un estilo de vida, tendiendo a evitar el sobrepeso, para lograrlo se debe incrementar la actividad física sostenida acompañado con una buena dieta.

En países occidentales donde se realiza tamizaje la mayoría de las pacientes con esta patología son asintomáticas(5). Los síntomas se manifiestan como descamación de la areola, formación de costra, rubor, formación de hoyuelos, retracción de los pezones, secreciones; aunque el más frecuente es la aparición de masas con bordes irregulares(5).

La detección temprana del cáncer de mama es de importancia ya que puede en algunos casos mejorar significativamente el pronóstico y las posibilidades de tratamiento(8). Para detectar esta patología en etapas asintomáticas la herramienta de menor costo efectividad es la mamografía 2D digital, que hoy se recomienda se emplee en un contexto de uso discriminado según riesgo (9). En este sentido, el Ministerio de Salud Pública de Uruguay produjo recientemente una nueva adaptación de guías internacionales(7). Considerando brevemente la guía, diremos que en ella se adopta un enfoque conforme al riesgo individual clasificándose en promedio, intermedio y alto, considerando factores genéticos, antecedentes personales y familiares como también densidad mamaria.

Estas recomendaciones están alineadas con la iniciativa mundial contra el cáncer de mama impulsada por la OMS y esta se estructura bajo tres ejes fundamentales, promoción para la detección temprana, diagnóstico oportuno y tratamiento integral.(10) El trabajo de Garau plantea un análisis del cáncer de mama en Uruguay teniendo en cuenta tres franjas etarias; 20 a 44, 45 a 69 y 70 años o más. Se evaluaron tendencias de incidencia, mortalidad, estadio al diagnóstico y perfil biológico tumoral en distintos intervalos de tiempo.(11) Como resultado se mostró una incidencia estable en la población general, pero con un notable aumento en el primer rango etario correspondiente a mujeres jóvenes, manteniendo una estabilidad en el grupo de 45 a 69 y disminución en mayores de 70 años. En cuanto a la mortalidad disminuyó,

principalmente en el segundo rango. La mayor cantidad de casos llegando a más de un 70% se detectan en estadios tempranos, siendo el subtipo predominante el luminal, mientras que en mujeres jóvenes es más frecuente diagnosticar tumores agresivos (HER2). En consecuencia, resulta fundamental implementar estrategias diferenciadas orientadas a la prevención, el diagnóstico y el tratamiento.(12)

Para garantizar una óptima calidad de imagen mamográfica, estas deben proporcionar el contraste adecuado entre los distintos tejidos, bajo ruido, baja dosis de radiación y una alta resolución espacial para observar elementos que pueden corresponder a lesiones malignas. Para conservar una calidad óptima se define un control de calidad que consta de un conjunto de pruebas, que ayudan a mantener estos estándares disminuyendo la probabilidad de que los valores salgan de tolerancia.(13) Dicho control de calidad implica cuatro pasos, pruebas realizadas para verificar el funcionamiento correcto de equipos nuevos o que han pasado por reparaciones significativas, establecer un rendimiento base del equipo, evaluación temprana de variaciones en el rendimiento del equipo, antes de que se manifiesten en las imágenes radiológicas y testeo del rendimiento del equipo tras un mantenimiento. No realizar los controles de calidad compromete la fidelidad de las imágenes diagnósticas, afectando principalmente al paciente, quien depende de resultados precisos para su atención médica.(13)

En este contexto, garantizar la calidad de las imágenes mamográficas resulta fundamental no solo para el diagnóstico precoz del cáncer de mama, sino también para la seguridad del paciente y la eficiencia de los servicios de imagenología. El método propuesto por el ACR ha demostrado ser útil para evaluar técnicamente el desempeño de los sistemas de mamografía; sin embargo, presenta limitaciones para valorar el rendimiento perceptivo de los observadores y, por ende, la calidad efectiva de las imágenes. Estas limitaciones son especialmente relevantes en la detección de microcalcificaciones, un proceso influenciado tanto por las propiedades físicas de la imagen como por factores visuales y cognitivos de los observadores. Dichos componentes pueden analizarse de manera diferenciada cuando participan múltiples lectores y se aplican modelos perceptuales adecuados(14). En este trabajo se propone y se evalúa preliminarmente un conjunto de nuevos métodos para la valoración de la calidad de imágenes de fantasmas mamográficos ACR, incorporando principios de la visión humana. El objetivo es avanzar hacia un enfoque híbrido que integre tanto la evaluación técnica como la perceptual, permitiendo una caracterización más completa y realista de la calidad de imagen.

Marco teórico

Estadificación del cáncer de mama. Para lograr una orientación adecuada sobre las decisiones terapéuticas del cáncer de mama, la estadificación inicial constituye un papel fundamental. Sin embargo, existen variaciones significativas en la práctica clínica, en cuanto a qué estudios solicitados y bajo qué criterios, lo que puede generar diferencias en la calidad asistencial.

El artículo Guidance for Canadian Breast Cancer Practice: National Consensus Recommendations for Clinical Staging of Patients Newly Diagnosed with Breast Cancer⁽¹⁵⁾ presenta un consenso nacional desarrollado por la Real Canadian Breast Cancer Alliance, cuyo objetivo es unificar criterios y establecer recomendaciones basadas en evidencia y opinión experta para la estadificación del cáncer de mama. Estas recomendaciones abarcan tanto la evaluación anatómica, como el uso de imágenes avanzadas y la marcación preoperatoria. Este propone que la indicación de estudios de estadificación para enfermedad a distancia debe basarse principalmente en el Grupo de Estadío Anatómico según la 8.ª edición del AJCC, evitando así el uso indiscriminado de técnicas de imagen en estadios tempranos. Los pacientes asintomáticos con enfermedad en estadios I o II (*Fig. 1*) generalmente no requieren estudios adicionales, dado que la evidencia internacional muestra bajo rendimiento diagnóstico en estos casos y un incremento en costos, retrasos terapéuticos. Por lo contrario para estadios más avanzados (IIB–IIIC), la realización de estudios como tomografía computarizada, gammagrafía ósea o PET-CT se considera pertinente debido al mayor riesgo de diseminación sistémica.

En cuanto al trabajo local y regional, el panel establece como estándar inicial la realización de mamografía bilateral y ecografía de mama y axila en todas las pacientes con sospecha de cáncer. Se recomienda considerar resonancia magnética en pacientes con antecedente de radioterapia torácica previa, dado su valor para diferenciar entre cambios por irradiación y enfermedad maligna

Las recomendaciones buscan promover una estadificación racional, basada en evidencia y adaptada al riesgo real de cada paciente, evitando tanto el sobrediagnóstico como la subestimación de la enfermedad. La iniciativa contribuye a la práctica clínica y ofrece un marco claro para la toma de decisiones, favoreciendo una atención más equitativa y consistente entre diferentes centros y regiones.

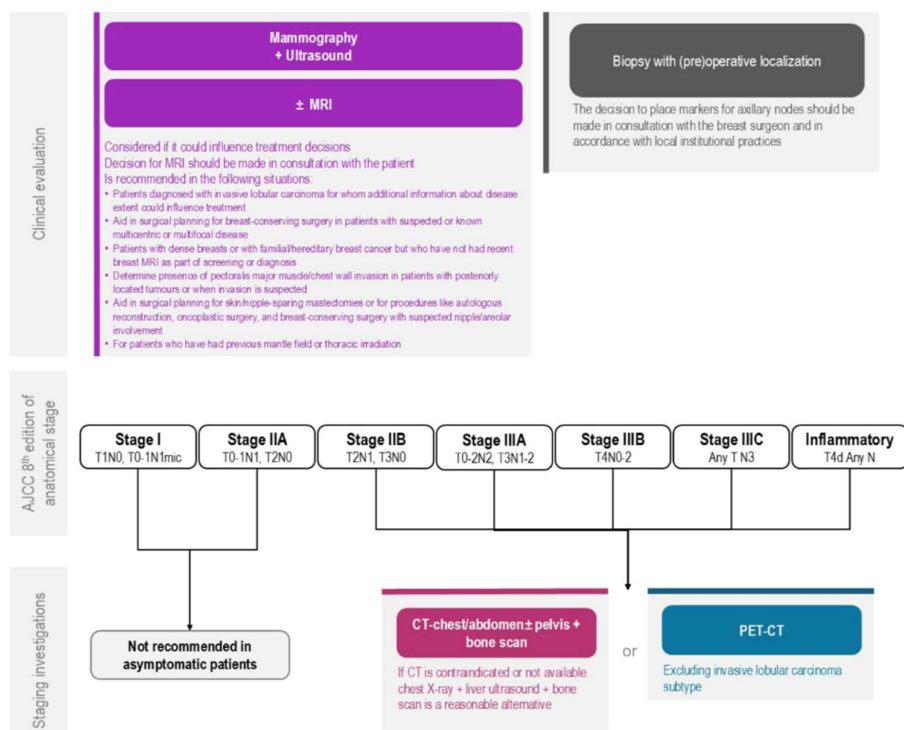


Fig. 1. Algoritmo recomendado para la estadificación inicial del cáncer de mama.

Avances del conocimiento y la atención clínica de mujeres con cáncer de mama.

Durante el siglo XIX hubo varias contribuciones a la investigación del cáncer de mama, una de las principales técnicas en este ámbito fue la introducida por el médico estadounidense William Halsted en 1890. La premisa central de Halsted sostenía que el cáncer se propaga radialmente desde el tumor primario hasta los ganglios linfáticos regionales, y sólo después a sitios distantes. Por lo que se consideraba al cáncer de mama una enfermedad local, posteriormente se concluyó que la mastectomía, la cual venía en auge, en ese entonces era la técnica con mayor probabilidad de disminuir la mortalidad en pacientes portadoras de esta patología. Halsted introdujo la mastectomía radical apoyada en bases científicas como tratamiento y erradicación del cáncer de mama, consistía en un amplio corte de piel, extirpación del pectoral mayor, disección axilar y extirpación en bloque de todos los tejidos, con un amplio margen, tratando de retirar el mayor tejido posible, tratando de evitar que se produjera una posible metástasis a distancia. Esta técnica se reforzó a través del atractivo de los cirujanos, ya que los convertía en el centro del tratamiento para dicho cáncer. Así mismo, este procedimiento se usó durante aproximadamente 100 años, siendo la cirugía predominante para el tratamiento del cáncer de mama.(3)

En el siglo XX emergieron nuevos métodos para el tratamiento contra el cáncer de mama, como quimioterapia y radioterapia, complementando el abordaje quirúrgico. Existieron cirujanos que cuestionaron esta premisa propuesta por Halsted, uno de ellos, el británico Geoffroy Keynes. Keynes desaprobaba totalmente la mastectomía radical y fue uno de los primeros en adentrarse en utilizar estos nuevos métodos, sugiriendo una cirugía más limitada y con seguimiento de radioterapia, comenzando a tratar pacientes con estadios avanzados, presentando algún éxito. Sugiriendo el cambio de la cirugía por estas metodologías menos

agresivas. Este cambio radical dejó la puerta abierta a que nuevos investigadores avancen en el tema, llegando a derrocar el modelo Halsted. En 1985 el cirujano estadounidense Bernard Fisher, en su estudio B-06 planteó que el cáncer era una enfermedad sistémica, en la que las células cancerosas podían viajar por el sistema circulatorio y linfático llegando a otras zonas del cuerpo. Aludió a que este tratamiento más conservador seguido por radioterapia, brindaba los mismos resultados en cuanto a sobrevida que la mastectomía radical, la teoría de Fisher respaldada por bases científicas atrajo al desarrollo y crecimiento de nuevas áreas como radioterapia, quimioterapia, radiología (mamografía), inmunoterapia, cirugía plástica, rehabilitación, psiquiatría, entre otras.(4)

El Dr. Fisher comenta. "Las estrategias terapéuticas para el manejo del cáncer de seno, han evolucionado en el tiempo, paso a paso, y son el resultado de información biológica, que ha llevado a un mejor entendimiento de la enfermedad. Es lógico anticiparse a que esto continuará, y los logros futuros se basarán en hipótesis biológicas nuevas. El cáncer de seno fue, es y será siempre relacionado a la Ciencia y no al populismo".(16)

Luego de múltiples trabajos realizados por científicos, se logra llegar a un hallazgo muy importante que cambia el transcurso de cómo se afrontaba el problema del cáncer de mama. El científico Axel Ullrich descubre la existencia del gen ERBB2, este gen se encontraba amplificado en el cáncer de mama. Al dirigir un anticuerpo al dominio extracelular de HER2, resultaba en la inhibición del crecimiento celular del cáncer de mama.

Conjuntamente la industria farmacéutica comenzaba un apogeo en cuanto a la fabricación y desarrollo de medicamentos dirigidos a dicha patología, centrando la investigación sobre dicho gen. La farmacéutica Genentech, propiedad del científico Axel Ullrich, pionera en fabricar el primer anticuerpo y posteriormente el primer fármaco "Herceptin" (Trastuzumab), numerosas investigaciones concluyen que el nivel de sobre vida aumentaba luego de aplicado, pero también apuntaban a largo plazo donde se observan graves efectos secundarios.(6) A lo largo del tiempo, se han desarrollado diversos fármacos para el tratamiento del cáncer.

En 2012, se publicaron los resultados del estudio "CLEOPATRA", una investigación financiada por Genentech y otras compañías farmacéuticas. Este estudio evaluó la combinación de tres medicamentos (Pertuzumab, Trastuzumab y Docetaxel) como tratamiento para el cáncer de mama. Los hallazgos de "CLEOPATRA" marcaron un hito en la historia de la oncología, ya que lograron registrar un aumento en la supervivencia de pacientes con cáncer de mama avanzado. Como resultado de estos avances, la FDA aprobó en 2012 el uso de Pertuzumab como parte del tratamiento para esta enfermedad.(4)

Teorías y paradigmas biomédicos sobre el cáncer de mama. Las teorías etiopatogénicas son aquellas que buscan explicar el origen de la enfermedad y cómo ésta se desarrolla para producir manifestaciones, desde las más frecuentes hasta las consecuencias más graves para la salud(17). La teoría de la prevención a través del tamizaje de enfermedades prevé que la detección de formas subclínicas de la enfermedad se relaciona con mejores resultados que la detección clínica, mejores resultados que incluyen mejoras en la calidad de vida y sobrevida.(18)

Las teorías diagnóstica y pronóstica proponen que existen procedimientos específicos que permiten clasificar, es decir que el caso en consideración pertenece a una población de casos ya conocida y que eso posibilita la predicción que corresponda(19,20). La teoría terapéutica explica que ciertas intervenciones causan cambios en la etiopatogenia de la enfermedad de modo que se producen estados de salud-enfermedad intervenidos con mejor calidad de vida

y supervivencia de las personas comparados con los estados de salud-enfermedad de las personas sin intervención(19). Las distintas teorías y modelos de la etiopatogenia, el tamizaje, el diagnóstico, pronóstico y el tratamiento del cáncer de mama se conjugan en formas coherentes particulares, que podríamos denominar como teorías o modelos combinados o mejor paradigmas biomédicos.

Las principales teorías etiopatogénicas sobre el cáncer de mama se centran en distintos aspectos. El origen y la diseminación del cáncer de mama son aspectos principales de las teorías de Virchow-Halsted, Grey-Auchincloss-Madden, de Keynes, de Fisher y de Tot. La primera postula al cáncer como una enfermedad local de la mama, sin que se precisen más detalles del origen, donde las consecuencias graves de la enfermedad derivan del avance local, regional, linfático y por último sistémico en una secuencia rígida:

Paradigma de Virchow-Halsted (o modelo anatomo-quirúrgico localista)(21,22)

Mama → Tumor en la mama → Músculos Pectorales → Ganglios Axilares → Metástasis a Distancia → Muerte

La mastectomía radical (resecar mama, músculos pectorales y ganglios linfáticos axilares evita/retarda metástasis y muertes).

No se conoce si el tamizaje es beneficioso.

Paradigma de Grey-Auchincloss-Madden (o modelo anatómico modificado)(23–25)

Mama → Tumor en la mama → Ganglios Axilares → Enfermedad Sistémica → Muerte

Mastectomía Radical Modificada (resecar mama y ganglios linfáticos axilares) evita/retarda metástasis y muerte (equivale en beneficios a la mastectomía radical).

Paradigma de Mustakallio-Fisher-Veronesi-Keynes (o modelo sistémico o biológico)(26–29)

Sector de la mama → Tumor en la mama y Enfermedad Sistémica → Muerte

La cirugía mamaria conservadora (cuadrantectomía), con resección ganglionar axilar y radioterapia evita/retarda metástasis y muerte (equivale en beneficios a la mastectomía radical modificada y a la mastectomía radical), sin embargo Fisher propuso como parte de modelo que el tratamiento local no afectaba la supervivencia.

Otros modelos incluyen:

Paradigma espectral (Hellman-Weichselbaum-Fisher)(6,16)

Mama → Tumor en la mama → Enfermedad Sistémica → Muerte

Paradigma de Tóth (o “sick lobe theory”, teoría del lóbulo enfermo)(30,31)

Lóbulo enfermo → Tumor o Tumores Multifoco/multicéntricos en la mama y Enfermedad Sistémica → Muerte

Paradigma “oligometastásico”(32–34)

Mama → Tumor en la mama y Oligometástasis → Enfermedad Sistémica → Muerte

“Modelo cáncer de mama oculto(35–37)

Mama → Metástasis/Oligometastasis/Enfermedad Sistémica → Muerte

Teorías sobre la visión humana. En la antigüedad se creía en dos teorías relacionadas con el concepto de visión humana, una proponía que el objeto emite alguna “efluencia” que

llegaba al ojo produciendo así la visión (teoría de la intromisión) y en contrapartida la teoría de la extramisión, la cual enuncia que el ojo enviaba un tipo de fuerza visual y está capturaba las propiedades del objeto. La antigua teoría de la intromisión dio origen a la actual teoría de la imagen retiniana, esta fue de gran éxito ya que tenía una gran potencia explicativa e integraba conocimientos físicos, fisiológicos y psicológicos llevando a la concepción de la percepción como un proceso constructivo(38).

Esta nueva teoría que establece que la percepción visual empieza con una imagen óptica formada en la retina fue aceptada ampliamente. En este contexto, los aportes de Curcio et al. en 1990 fueron fundamentales (39). Curcio y cols. demostraron que: A) La distribución de los fotorreceptores no es homogénea sino que presenta una organización espacial heterogénea y precisa; B) La topografía de estos fotorreceptores (conos y bastones) evidencia diferencias entre los ejes horizontal y vertical (*Fig. 2*); C) Los conos, que brindan soporte a la visión diurna detallada y colorida, se distribuyen con una franja de alta densidad que se extiende en el meridiano horizontal, mientras que en el eje vertical se observó una densidad ligeramente mayor sobre la retina inferior con respecto a la superior; D) La fóvea presenta contornos levemente elípticos con mayor extensión en el eje horizontal, optimizando la organización anatómica en el eje horizontal(40). Estas asimetrías causan que el rendimiento visual humano varía a lo largo del campo visual y variando incluso en ubicaciones isocinéticas, anisotropía horizontal-vertical y una simetría en el meridiano vertical, desencadenando un mejor rendimiento en el meridiano horizontal que en el vertical, y menor en la región superior que en la inferior del meridiano vertical.(40)

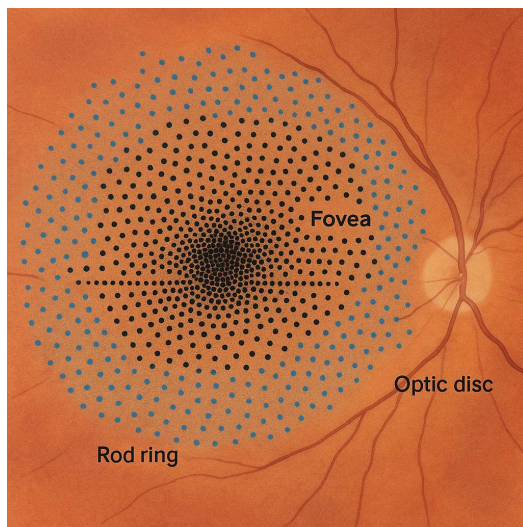


Fig. 2. Distribución topográfica de fotorreceptores en la retina humana. Esta muestra la distribución topográfica heterogénea de los fotorreceptores descrita por Curcio et al., mostrando la alta densidad de conos en la fóvea de contorno ligeramente elíptico y extendido en el eje horizontal y el anillo perifoveal de bastones que caracteriza la organización espacial de la retina.

Desde otros distintos enfoques se han aportado marcos relevantes en el área de la interpretación de imágenes médicas como lo es la teoría constructivista de Gregory y

Helmholtz que sustenta que la percepción no depende únicamente de información sensorial, sino que tiene en cuenta procesos cognitivos que integran la experiencia previa y las expectativas del observador, según esta teoría la mente interpreta los estímulos mediante inferencia inconsciente asociando información aprendida entre determinadas sensaciones y los objetos que las producen. Lo percibido depende de la estimulación actual como de experiencias construidas desde el pasado(41). Sin embargo, esto puede no ser tan relevante si pensamos para la tarea psicofísica que se realiza al informar sobre la detección o no de imágenes de microcalcificaciones simuladas en fantasmas mamarios. En tal caso, pensamos que debe predominar la agudeza o salud sensorial, en este caso visual. De hecho, Gibson propuso una teoría más simple, denominada como teoría ecológica, que plantea que la percepción se enmarca (o se puede enmarcar) directamente en la información visual disponible sin la necesidad de inferencias cognitivas complejas(42). En esta teoría la calidad técnica de las imágenes es de suma importancia ya que el sistema visual utiliza información disponible sin procesos compensatorios.

Es importante mencionar que la detección de imágenes de microcalcificaciones simuladas por los observadores humanos requiere la búsqueda e identificación (o no identificación) de las señales correspondientes (ver esquema de abajo). En los estudios psicofísicos actuales se proponen pruebas donde estos procesos ocurren en un tiempo limitado, menor a 300 ms(43)

En las pruebas de percepción visual la mirada se fija al comienzo en un punto fijo central de la pantalla rectangular con un tiempo amplio, luego se proporcionan estímulos y se recogen respuestas con tiempos regulados y acotados. Se suelen realizar dos tipos de pruebas: las de búsqueda(42,44) y las de detección, entre estas últimas se suelen realizar con fondos naturales (como en la tesis de Zhang(43)) o sobre fondos con ruido blanco. En las pruebas de detección, las señales están presentes en el 50% de los estímulos, los que se le presentan a los observadores en una secuencia aleatoria. Los observadores en estas pruebas psicofísicas saben donde debería estar la señal (en el centro del círculo), por tanto la búsqueda o fijación de la mirada es relativamente rápida. En los otros ensayos, los de localización, la señal está en todos los estímulos pero en distintas localizaciones posibles dentro del círculo o zona de búsqueda). Es importante señalar que estas pruebas están diseñadas para evaluar las capacidades perceptuales de los observadores. En cambio, las pruebas realizadas sobre imágenes de fantasmas tienen la particularidad de no tener reglas de tiempo ni de fijación inicial de la mirada y el objetivo original propuesto es decidir si acreditar o no un servicio mamográfico y como ya señalamos, en nuestro caso el objetivo es estimar la calidad de las imágenes y decidir que método de medición de la calidad de las imágenes es mejor.

Teorías sobre la formación de imágenes de rx (sistemas lineales). Al estudiar cómo se forma una imagen de rayos X, una de las formas más útiles de describir el sistema es considerarlo como un sistema lineal. Esta aproximación permite describir el comportamiento del sistema de manera proporcional y predecible, facilitando el análisis de cómo la señal de entrada (rayos X) se transforma en la señal de salida (imagen).

La lógica exigiría que definamos el término “sistema” antes de intentar definir un “sistema lineal”, entonces nos concentramos en que un sistema es un operador que transforma una señal de entrada en una señal de salida mediante una relación bien definida.

En matemáticas y procesamiento de señales:

$$y = S\{x\}$$

en donde:

x = señal de entrada (fantoma),

S = el sistema (la operación),

y = señal de salida (juicio humano). (45)

La Teoría de Sistemas Lineales es una herramienta matemática que se usa para modelar y analizar sistemas que asumimos cumplen dos propiedades clave como la linealidad y la invariancia en el tiempo o espacio(46). En radiología, esta teoría se aplica en varios aspectos, sobre todo en el procesamiento y formación de imágenes médicas, lo que para este Trabajo Final de Grado es muy relevante, por ello, en lo que sigue tomaremos como referencia principal el libro "Radiological imaging: the theory of image formation, detection, and processing" de Barrett y Swindell(pag. 29 en (46)).

En un sistema de rayos X, podríamos llamar al coeficiente de absorción de rayos X en algún punto "r" del paciente como la entrada "x" y a la densidad óptica de la película revelada la salida "y". Asumiendo esa definición, que puede ser adecuada para cierto propósito de análisis o estudio, el sistema es todo aquello que explica la transformación de "x" en "y". Alternativamente, la atenuación total de un haz de rayos X al atravesar una trayectoria particular a través del paciente podría considerarse la entrada "x", y el flujo de fotones de rayos X en una ubicación particular de la película podría ser la salida "y". En estos ejemplos, la señal de entrada así como la de salida están asociadas a la variable independiente espaciotemporal, puede ser una posición espacial uni, bi o tridimensional asociada a la variable tiempo. Lo anterior es apropiado para tratamientos clásicos en coordenadas cartesianas de los sistemas.

Ahora es sencillo definir el concepto de linealidad, supongamos que una entrada "w(u)" produce la salida "wo(u)" y que una entrada diferente "w2(u)" produce la salida "wo(u)". Se dice que el sistema es lineal si, cuando ambas entradas se aplican juntas de modo que:

$$w(u) = w(u) + w2(u)$$

la salida viene dada por:

$$Wout(u) = wo(u) + w2(u).$$

Es importante tener en cuenta que la linealidad es una idealización y ningún sistema físico real es estrictamente lineal. De hecho, aunque el ennegrecimiento de una película no es proporcional al flujo de fotones incidentes, sin embargo, la teoría de sistemas lineales, es una herramienta extremadamente poderosa en muchas circunstancias. Los efectos de las no linealidades omnipresentes a menudo pueden considerarse por separado e incorporarse a la teoría lineal casi como una ocurrencia posterior(pag. 31 en (46))

Para continuar con el tema, Barrett y Swindell proponen considerar una cámara simple en la que se proyecta una imagen de un objeto plano con un aumento unitario sobre una película plana, la variable "u" es entonces un vector bidimensional en el plano del objeto o en el plano de la imagen, sin embargo, la irradiancia en algún punto de la imagen no está determinada únicamente por la luz emitida desde el punto correspondiente en el plano del objeto. En

cambio, la lente produce inevitablemente una imagen más o menos borrosa del punto del objeto, entonces podemos pensar que cada punto del objeto contribuye a la respuesta en muchos puntos de la imagen, o equivalentemente, cada punto de la imagen recibe un estímulo de muchos puntos del objeto. Este tipo de comportamiento no local se puede describir de forma bastante general para sistemas lineales escribiendo "wout(u)" como una superposición lineal de los valores de "win" para todos los valores de "u". Dado que "u" suele ser una variable continua, la superposición lineal más general es una integral de la forma (pag. 31 en (46)).

En imágenes, la convolución es una operación que se emplea para representar cómo el sistema "difumina" o altera la señal original, usándose por ejemplo para modelar el desenfoque o para aplicar filtros de mejora de imagen (pag. 36 en (46)).

En la formación de imágenes técnicas como radiología convencional, tomografía computada (CT) y resonancia magnética (MRI), el cuerpo se modela como un sistema que transforma un estímulo (haz de rayos, campo magnético, etc.) en una imagen. El sistema imagenológico puede considerarse como un sistema lineal que actúa sobre una función de entrada (la anatomía) para producir una salida (la imagen). Cada punto del cuerpo produce una pequeña "respuesta" en la imagen, esta respuesta se llama función de dispersión del punto (en inglés: Point Spread Function, PSF) (pag. 190 en (46)).

En un sistema lineal, la imagen completa se puede entender como la convolución entre el objeto y la PSF del sistema. Entonces supongamos que el cuerpo es el objeto $f(x)$, y la imagen final es $g(x)$. Si el sistema es lineal e invariante, entonces:

$$g(x) = f(x) * h(x)$$

Donde $h(x)$ es la función de respuesta del sistema (PSF) y $*$ representa la operación de convolución (pag. 34 en (46)).

El análisis en el dominio de la frecuencia (por transformada de Fourier) permite estudiar cómo el sistema responde a diferentes detalles (frecuencias espaciales). Esto lleva al concepto de MTF (en inglés: Modulation Transfer Function), que describe qué tan bien el sistema reproduce distintos niveles de detalle. La mayoría de los sistemas lineales en general, y la mayoría de los sistemas de imágenes en particular, degradan la señal de entrada de alguna manera, por ejemplo, la MTF de la mayoría de los sistemas es menor que la unidad en todas las frecuencias (excepto cero) y disminuye rápidamente en las frecuencias más altas. Esto significa que los componentes de alta frecuencia, a menudo las características más importantes de la señal de entrada, tienen un contraste reducido en la salida (pag. 42 en (46)).

Una segunda fuente importante de degradación es el ruido, en cierto modo, nuestro tratamiento de los sistemas lineales hasta este punto ha sido "fraudulento" porque implica que existe una correspondencia única entre la salida y la entrada. Debido a estas limitaciones de MTF y ruido, un sistema lineal puede no cumplir su objetivo. Si se trata de un sistema de imágenes radiológicas, la imagen de salida puede no mostrar el detalle necesario para un diagnóstico preciso, en este caso, a menudo se puede obtener un rendimiento mejorado introduciendo la salida del sistema en un segundo sistema lineal llamado filtro. La función del filtro es para deshacer parte de la degradación introducida por el sistema primario. Por supuesto, la distinción entre el sistema primario y el filtro subsiguiente es meramente

semántica (diferencia terminológica). Podríamos igualmente llamar filtro al sistema primario o a cualquier sistema lineal.

En mamografía 2D se aplica la teoría de sistemas lineales. Esta teoría es muy útil para evaluar, optimizar y entender la calidad de imagen, especialmente porque se trata de detectar estructuras muy pequeñas como microcalcificaciones ya que necesitamos una imagen con altísima resolución espacial y bajo ruido para detectar lesiones sutiles. La teoría de sistemas lineales permite entonces modelar el sistema imagenológico, cuantificar su capacidad para reproducir detalles, optimizar parámetros como el detector, filtro, y compresión, evaluar calidad de imagen y dosimetría. Todo el sistema de mamografía (fuente de rayos X, el seno, el detector) se modela como un sistema lineal e invariante en el espacio.

La MTF evalúa qué tan bien se reproducen los diferentes detalles (frecuencias espaciales), debido a que en mamografía, se requiere una alta MTF a altas frecuencias, porque las microcalcificaciones son detalles muy pequeños. En la práctica, hay componentes no lineales, como la compresión del seno, procesamiento digital no lineal (realce, compresión de contraste), la respuesta del detector en saturación, aún así, la teoría de sistemas lineales es muy útil como modelo base para diseño y evaluación. Entonces como conclusión la mamografía 2D puede modelarse como un sistema lineal e invariante, siendo la imagen final la convolución del objeto con la respuesta al impulso (PSF) del sistema, usando la MTF para evaluar resolución espacial, permite comparar equipos y optimizar calidad vs. dosis.(10)

Paradigmas de evaluación de la exactitud de los juicios de los observadores sobre la detección y posición de microcalcificaciones en fantasmas mamográficos ACR. La evaluación de la exactitud de los juicios de los observadores centrado en la detección y posición de señales de microcalcificaciones se basa en los principios de la Teoría de Detección de Señales (en inglés: Signal Detection Theory, SDT). Anteriormente a esta teoría, el principal modelo psicofísico era la Ley de Weber-Fechner(47)

$$dp = k * (dS / S) \quad (\text{ec.1})$$

Esta representa únicamente cómo varían la percepción (dp) y el estímulo físico (dS) de manera local. Para comprender la relación global entre ambos, y poder introducir el concepto de umbral perceptual, es necesario considerar también su forma integral. Al integrar la ecuación diferencial se obtiene la expresión general propuesta por Fechner:

$$p = k * \ln(S / S_0) \quad (\text{ec.2})$$

Donde S_0 corresponde al umbral absoluto del estímulo. En esta formulación, la percepción es proporcional al logaritmo neperiano de la intensidad del estímulo, estableciendo que existe un umbral fijo a partir del cual dicho estímulo es percibido. De hecho, fue Fechner, en 1860, quien otorgó la forma final a este modelo psicofísico(12).

Posteriormente, en la década de 1950, Tanner y Swets reformularon el marco conceptual mediante el desarrollo de la Teoría de Detección de Señales(13). Este enfoque introdujo una descripción cuantitativa del proceso por el cual los observadores humanos toman decisiones visuales bajo condiciones de incertidumbre, reemplazando la noción clásica de un “umbral de detección” rígido por un modelo estadístico en el cual la tarea perceptual consiste en discriminar entre señal y ruido. Según estos autores, la detección depende de la

toma de decisiones basada en la evidencia sensorial disponible, la cual se encuentra inevitablemente afectada por múltiples fuentes de ruido, tanto internas como externas.(13) En consecuencia, la respuesta del observador no refleja únicamente la presencia física del estímulo, sino también la forma en que éste integra y evalúa dicha evidencia.

Este modelo representa la actividad perceptual mediante dos distribuciones de probabilidad: una correspondiente al ruido y otra a la combinación de señal más ruido. La separación entre ambas distribuciones cuantifica la sensibilidad del observador, expresada mediante el parámetro d' , que refleja su capacidad para discriminar la presencia de una señal en condiciones de incertidumbre.

Otro parámetro de suma importancia es el criterio de decisión (β) el cual refleja el sesgo o estrategia con la que el observador elige si una señal está presente o no. Permitiendo analizar por separado la sensibilidad perceptiva del observador y su criterio decisional, siendo estos esenciales en la evaluación de exactitud de los juicios radiológicos.(48) Es importante notar que la afirmación anterior es cierta solo en el contexto de pruebas donde quien diseña o pone la prueba conoce qué respuestas son correctas, es decir, que podemos analizar los juicios de los evaluadores de las imágenes (que en realidad más que evaluadores son los evaluados) sobre las imágenes sabiendo si estaban justificados o no. Como ya deslizamos en el comentario anterior, esta es una teoría que intenta explicar los juicios humanos o las decisiones de cualquier sistema de decisión. No hay en principio nada en la teoría que no se pueda aplicar a un algoritmo de aprendizaje automático y visión por computadora. Distinto es cuando estamos trabajando con imágenes de fantasmas (donde la señal siempre está y el evaluador es ambas cosas, evaluador y evaluado. Normalmente, el ámbito de la visualización y detección de microcalcificaciones en imágenes de fantasmas mamográficos es un desafío ya que se discrimina señales débiles en un entorno de ruido generado por el mismo fantoma o por variaciones de dosis. Por otro lado, la evaluación de los observadores requiere considerar conjuntamente la incertidumbre visual (es decir el ruido perceptual del observador, no de la imagen), la decisión estadística (criterio decisional interno de cada observador, que tampoco está relacionado a la física del fantoma) y la variabilidad interindividual de estos factores.(14)

Esta teoría da origen ciertos paradigmas de evaluación de la exactitud unos los principales son el análisis ROC (Receiver Operating Characteristic) y su extensión el FROC (Free-response ROC), que permiten cuantificar la relación entre la tasa de verdaderos positivos y la de falsos positivos mediante diversos criterios decisionales.(14,49) Pudiendo medir tanto la capacidad de detección como la precisión de localización, obteniendo datos que permiten evaluar la calidad de las imágenes y el rendimiento perceptivo de los observadores(49,50). Este método ROC anteriormente mencionado ha sido utilizado frecuentemente para evaluar rendimiento de los sistemas de imágenes y/o de observadores al representar gráficamente la relación entre la tasa de verdaderos positivos (sensibilidad) y la tasa de falsos positivos (1- especificidad) bajo distintos umbrales de decisión. La construcción de este gráfico ROC se calcula para cada posible umbral de decisión los valores de sensibilidad y 1- especificidad representandose en un plano cartesiano, esta curva muestra el comportamiento del modelo o del observador frente a diferentes criterios de decisión, pudiendo cuantificar la capacidad de distinción entre verdaderos y falsos estímulos. El área debajo de esta curva (AUC) indica el rendimiento global, un valor cercano a 1 muestra una alta capacidad discriminativa y valores próximos brinda una

capacidad similar al azar. Esta utilización muy frecuente no es aplicable cuando la tarea de decisión implica más que una determinación de presencia o ausencia.(51) A menudo se necesita localizar regiones específicas de la imagen como puede ser una microcalcificación en una imagen de fantoma mamográfico, esta información adicional de localización no se puede analizar mediante este método ROC llevando a una pérdida de potencia estadística. Para contrarrestar esta limitación se desarrolló el paradigma FROC (Free-Response ROC) el cual permite múltiples marcas de ubicaciones al observador en una misma imagen, concediendo el análisis simultáneo de detección y localización. Cuando esta marca se encuentra dentro de un radio de aceptación predefinido se considera una verdadero positivo, cuando no corresponde a ninguna señal se considera falso positivo(52,53). A partir de estas clasificaciones se construye la curva FROC, que representa la tasa de verdaderos positivos de localización frente a la tasa de falsos positivos por imagen(49,50). FROC es útil cuando el observador produce múltiples respuestas por imagen, como en el caso de detección de microcalcificaciones en imágenes de fantasmas mamográficos, este permite calcular un índice denominado Figura de Mérito (FOM), análogo al área bajo la curva del método ROC, FOM indica la capacidad conjunta de decisión y localización de los observadores(53,54).

Teorías sobre la detección de microcalcificaciones en imágenes 2D mamográficas.

La detección de microcalcificaciones es importante ya que aproximadamente el 40% de los casos de cáncer de mama llevan consigo la aparición de estas según lo reportado por Mandelson et al. (2000) y Duffy et al. (2019) (55,56). De hecho, las microcalcificaciones son un biomarcador del Carcinoma Ductal In Situ (DCIS por su sigla en inglés), el cual es una lesión tumoral benigna que se asocia en una proporción de los casos a malignidad(57). Estas están formadas por acumulaciones de calcio en las células mamarias, produciendo la máxima atenuación de los rayos X. La dificultad de detección de las microcalcificaciones depende en gran medida de la densidad de los tejidos adyacentes, en mamas con abundante tejido fibroglandular la detección se dificulta, mayoritariamente sucediendo en mamás jóvenes, lo que puede desencadenar un problema ya que la evolución de la enfermedad puede llegar a ser más rápida. La detección se encuentra influenciada por factores humanos, estos son ajenos a las características de las imágenes(54). Elizabeth Krupinski(58) escribió alguno de los principales factores, como son la experiencia y entrenamiento, mediante pruebas estadísticas llego a la conclusión de que los observadores con mayor experiencia necesitaban menos tiempo para la visualización ya que fijaban la atención muy rápidamente hacia la ubicación de la lesión cuando se presentaba una imagen, mientras que a los observadores inexpertos les llevaba en promedio el doble de tiempo para ubicar esta primera lesión(58). A su vez los observadores expertos lograron disminuir la proporción de falsos positivos(58).

Estas diferencias entre observadores expertos e inexpertos no solo están adjudicadas por la experiencia y el entrenamiento, sino también pueden ser explicadas parcialmente desde la perspectiva más simple de modelos psicofísicos que ya introdujimos porque ponderan la relación entre estímulos físicos y cambios en la percepción visual (14,59). Una de las principales herramientas es la función psicométrica la cual modela la relación entre el estímulo físico y la respuesta del observador, describiendo la probabilidad de detección de este estímulo en función de una variable física (tamaño de microcalcificación). Esta función es del tipo sigmoidea describiendo una transición progresiva entre dos estados mayoritariamente, entre respuestas correctas e incorrectas a medida que la intensidad del

estímulo aumenta(14). Se utiliza muy a menudo en pruebas para evaluar la capacidad de un observador para detectar o discriminar estímulos ya que representa a la perfección una transición desde el nivel de azar hasta un rendimiento cercano al cien por ciento, refleja la probabilidad de la percepción visual (55). Por su parte, la toma de decisiones visuales influye significativamente en la interpretación y detección frente a imágenes mamográficas. Esta toma de decisión se puede explicar al menos parcialmente mediante el modelo "Signal Detection Theory" considera que esta se toma bajo condiciones de incertidumbre evaluando la presencia o ausencia de microcalcificaciones, este desempeño se analiza mediante parámetros como son la sensibilidad (verdaderos positivos) y especificidad (verdadero negativos). La autora Elizabeth Krupinski determinó que las decisión de verdaderos positivos está relacionada con tiempos de visualización más largos mientras que los verdaderos negativos se asocian con tiempos más cortos, esto confirma que el tiempo es un predictor útil de la calidad de los informes médicos.(58)

Al incorporar nuevas tecnologías en los sistemas digitales de mamografías, se ha aumentado la detección al mejorar la resolución y el procesamiento de las imágenes. Por su parte se han desarrollado herramientas que influyen en la toma de decisiones por parte del radiólogo, estas van desde simples algoritmos hasta otros más complejos que detectan microcalcificaciones en mamografías en forma automática. En la visualización de imágenes mamográficas la utilización de software complementarios es de gran utilidad, ya que debido a la complejidad de la arquitectura mamaria y el gran número de imágenes analizadas es muy común la distracción de radiólogos. Uno de los recursos digitales más utilizados es el sistema Computer Aided Diagnosis (CAD), siendo una red neuronal que emplea algoritmos para que resalten áreas sospechosas, el propósito esencial de esta es ofrecer una evidencia más objetiva aumentando la confianza diagnóstica del radiólogo, llevando a mamografías con una disminución en la cuantificación de falsos negativos(34). Estos últimos años se ha implementado la inteligencia artificial en estos temas, como concluye Rodriguez Alejandro los resultados mostraron claramente se ha reducido la disparidad entre estos sistemas y los expertos humanos aunque la IA sigue siendo inferior a los mejores radiólogos. Como el sistema antes mencionado se puede dar una utilidad como ayuda al experto apoyando a la toma de decisión, marcando posibles lesiones y advirtiendo errores de interpretación y descuidos de lectura. Aunque representa un sistema sumamente prometedor, su implementación demanda aún una investigación rigurosa y sistemática.(35)

Teorías sobre la información y la pérdida de exactitud. La Teoría de la Información sostiene que todo método de medición puede concebirse como un sistema que transmite información desde un fenómeno físico hacia una representación o estimación de ese fenómeno. Claude Shannon sostuvo que, la cantidad de información que se conserva determina la precisión del método, el método más exacto es aquel que mantiene la mayor información sobre la señal original, esto reduce la incertidumbre asociada a la medición. Cuando el proceso de observación introduce simplificaciones ya sea por limitaciones perceptuales o conceptuales, se produce una pérdida de información, lo que provoca un aumento de la entropía del sistema lo que disminuye la exactitud del resultado obtenido. Volviendo al caso en cuestión, el tamizaje del cáncer de mama podemos decir y quizás asumir

que en la mama de la paciente con cáncer subclínico está toda la información necesaria para especificar ese diagnóstico y más aún. Esto es en la gran mayoría de los casos así. Cualquier método de tamizaje brindará información incompleta y sesgada de esa realidad física. (61)

La Teoría de la Información relaciona directamente la exactitud (en la literatura ingenieril muchas veces se usa precisión como sinónimo de exactitud, en bioestadística y medicina se prefieren usar esos dos términos con significados distintos) y la cantidad de información, cuanto mayor es la información disponible de un fenómeno menor es la incertidumbre lo que aumenta la capacidad de estimar las características de este fenómeno correctamente.(61) A nivel metodológico, lo enunciado por Luce y Tukey sostiene que los métodos más simples y con menor información pueden derivarse de métodos con mayor cantidad de información mediante procesos de reducción o categorización, estos procesos llevan a una pérdida de información aunque mejora la aplicabilidad o la facilidad interpretativa.(37) En el ámbito de la detección visual en imágenes mamográficas de fantasmas ACR como es el caso de nuestra investigación, estas teorías permiten comprender y explicar cómo los distintos métodos de evaluación implican diferentes niveles de información lo que lleva a distintos grados de exactitud. Los métodos cuantitativos continuos como la medición del error posicional permiten tener una mayor cantidad de información sobre la ubicación, lo que ofrece una descripción más detallada, en cambio métodos discretos que clasifican “se ve/no se ve” o la escala de puntuación como ACR simplifican la información visual facilitando el análisis y la comparación pero perdiendo información del estímulo perceptual que el observador puede evaluar.

Antecedentes

Los principales antecedentes para esta investigación son el Método del Colegio Americano de Radiología (en inglés, American College of Radiology, ACR) para la acreditación de servicios, el método de Huda para medir la calidad de las imágenes de fantasmas ACR (2002) y el método automático propuesto por Sundell que también usa fantasmas del ACR (2018).(63)

Método del ACR. El ACR creó en 1987 un método para el control de calidad en sistemas de imagen el cual está constituido por un conjunto de dispositivos físicos, protocolos y pruebas destinadas a evaluar el desempeño técnico de los mismos. Su principal objetivo es que se mantenga un nivel de calidad constante que permita detección de lesiones sutiles como fibras, masas y microcalcificaciones. Pondremos foco en estas últimas, ya que como señalamos es uno de los principales indicadores de lesiones que con una probabilidad relevante pueden anteceder o coexistir con el cáncer de mama, como ya mencionamos. El método requiere la concurrencia interdependiente del radiólogo, del tecnólogo y del físico médico, cada uno tiene responsabilidades específicas. Una de las pruebas más significativas del método es el análisis del fantoma ACR, que permite evaluar la capacidad del sistema para reproducir objetos de bajo contraste y pequeño tamaño. Para ello el tecnólogo o el físico médico observa la imagen obtenida cuantifica qué cantidad de lesiones simuladas son visibles y calcula una puntuación para la imagen. El control requiere la realización de revisiones de componentes del sistema como el buen funcionamiento de la fuente de RX, los monitores de visualización, el buen funcionamiento del brazo y paleta de compresión, evaluación de la dosis, la limpieza del sistema, etc; buscando garantizar una calidad de imagen homogénea minimizando errores de interpretación.(1)

El fantoma con el cual se realiza el método antes mencionado en equipos de mamografía digital 2D indirectos corresponde al fantoma mamográfico de acreditación modelo 015 Computerized Imaging Reference Systems (CIRS) este dispositivo está estandarizado para control de calidad por parte del American College of Radiology (ACR). El fantoma intenta simular una mama comprimida de 4,2 cm de espesor componiéndose por una base acrílica, una cubierta transparente y un bloque de cera donde están contenidos los objetos de prueba, siendo un total de dieciséis juegos objetos; compuesto por seis fibras formadas por nylon (1,56;1,12;0,89;0,75;0,54,0,40 mm), cinco masas (2,00;1,00;0,75;0,50,0,25 mm) y cinco grupos con 6 microcalcificaciones en cada uno de ellos donde su material para desarrollarlas es óxido de aluminio (Al_2O_3), siendo estas de distintos tamaños (0,54;0,40;0,32;0,24;0,16 mm). Para evaluar el umbral de detección del sistema y evaluar la resolución espacial junto con la sensibilidad de contraste, dichas estructuras se distribuyen en decreciente tamaño y contraste.(64)

La puntuación de los distintos tipos de objetos que simulan lesiones se asigna según su detección o no detección en la imagen del fantoma ACR. A las fibras se les asigna una puntuación individual, si se visualizan de 8 mm o más de longitud en su ubicación y orientación correcta y se asigna ½ punto cuando la longitud es mayor a 5mm y menor a 8mm. A las masas se le asigna 1 punto si el estímulo es visible en su correcta ubicación y esta se visualiza circular, cuando esta masa se encuentra en su correcta localización pero no se observa circular se asigna ½ punto. Por último, a los grupos de microcalcificaciones se le asigna 1 punto cuando se visualizan de 4 a 6 microcalcificaciones en su localización correcta y ½ punto si se visualizan 2 o 3 en su ubicación correcta. Estas puntuaciones se suman alcanzando un puntaje máximo global de 16 puntos. Para que el sistema de mamografía se encuentre dentro de los límites de calidad aceptable y el servicio pueda ser acreditado según el estándar de ACR se deben visualizar al menos tres grupos de microcalcificaciones claramente visibles sin artefactos junto con cuatro fibras y tres masas, en el caso que la cantidad sea inferior a estos valores mínimos el equipo no aprueba el examen de calidad y requiere revisión técnica.(1)

“Método de Huda”: Huda(2002) junto a otros autores importantes como Dance, en su estudio tuvo como objetivo principal evaluar la eficacia del fantoma de acreditación del ACR para valorar la calidad de imagen en sistema mamógrafos digitales(2). En su investigación los autores generaron imágenes del fantoma ACR variando distintos parámetros de adquisición como el mAs y el kVp y las imágenes fueron visualizadas por 8 observadores de distintas profesiones entre ellos físicos médicos, radiólogos, residentes y tecnólogos. Los observadores evaluaron la cantidad de objetos visibles en las imágenes, siendo estos como ya dijimos fibras, microcalcificaciones y masas. Los resultados obtenidos mostraron que al aumentar el mAs la visibilidad de estos objetos aumentó progresivamente, mientras que al aumentar el kVp superando los 26 kVp el número de objetos se mantuvo constante, y al reducir este número a 24 kVp disminuyó el número de objetos detectados, también se observó que el desempeño de los observadores se estabilizó cuando la dosis glandular promedio superó 1,6 mGy lo que indica que la detectabilidad de las lesiones simuladas se produce en un rango de dosis menor al empleado en la práctica clínica habitual (con servicios que funcionan correctamente).

A partir de estos resultados obtenidos los autores concluyeron que el diseño actual del fantoma ACR no resulta satisfactorio para evaluar la calidad de imagen en mamografía digital.(2)

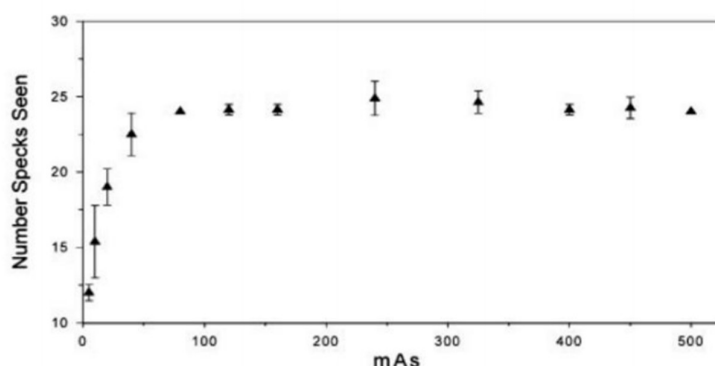


Fig. 3. Cantidad absoluta de microcalcificaciones visible por imagen en función del mAs.

En la (Fig. 3) del estudio, se observa la cantidad absoluta de microcalcificaciones visible por imagen en función del mAs y cada dato es el valor medio de los ocho observadores. Podemos observar en una primera parte de la curva un aumento esperado teóricamente donde al incrementar la dosis aumenta la cantidad de microcalcificaciones esperadas, pero luego se puede observar una meseta donde al aumentar esta dosis no aumenta la visibilidad de las microcalcificaciones no siguiendo la teoría esperada. Esta investigación realizada por Walter Huda es de gran importancia, para el trabajo presentado ya que aborda paradigmas de detección para discriminar señales de bajo contraste como las manifestadas por las microcalcificaciones. También evidencia la necesidad de replantear métodos de evaluación de calidad de imagen basado en fantasmas ya que no representan las condiciones perceptuales del observador en sistemas digitales. En estas circunstancias el artículo tenido en cuenta mantiene que es importante utilizar modelos psicofísicos y herramientas tales como análisis ROC o FROC, ya que estos cuantifican la relación entre calidad física y la percepción humana.

Los resultados de Huda et al. indican que el fantoma ACR no fue concebido para evaluar de manera cuantitativa la calidad de imagen ni para operar en el rango de parámetros donde los sistemas digitales exhiben su desempeño nominal(2). En particular, estos autores demostraron que la detectabilidad de los objetos simulados (fibras, masas y microcalcificaciones) mejora con el incremento de la dosis hasta aproximadamente 1,1–1,6 mGy, pero se estabiliza por encima de este nivel, el cual coincide con las dosis clínicas usuales ($\geq 1,6$ mGy). En consecuencia, el fantoma deja de discriminar diferencias entre técnicas o configuraciones del sistema que funcionan dentro de parámetros operativos correctos.(2)

Discrepamos, no obstante, con la inferencia de Huda et al. de que el fantoma ACR sería inapropiado para propósitos de acreditación. De hecho, sus hallazgos son congruentes con el objetivo explícito del programa ACR: establecer un umbral mínimo de rendimiento y verificar que el sistema no haya experimentado una degradación significativa. Desde esta perspectiva, el fantoma ACR cumple adecuadamente su función como prueba de umbral y no debe interpretarse como un instrumento diseñado para comparar o optimizar sistemas de mamografía digital. Finalmente, reconocemos que el método empleado por Huda et al.

tampoco constituye la medida más exacta de la calidad de imagen, dado que reduce la evaluación al conteo de objetos simulados detectables, ignorando otras características relevantes de la imagen. Este conjunto de consideraciones orientó la formulación de las hipótesis principales de este estudio.

“Método de Sundell”: Sundell et al. busca desarrollar y validar un sistema automatizado para evaluar la calidad de imágenes mamográficas mediante la utilización del fantoma de acreditación ACR. Los autores para ello realizaron una metodología la cual intenta sustituir la evaluación visual tradicional, por un análisis mediante el uso de algoritmos de aprendizaje automático. Para su validación se usaron sesenta imágenes del fantoma ACR adquiridas por distintos equipos y adicionalmente veinte imágenes adquiridas con diversos niveles de dosis glandular. Los resultados generales mostraron una alta concordancia entre el análisis automático y la evaluación visual por observadores para fibras, mientras para microcalcificaciones y masas el sistema automático mostró una sensibilidad superior. Lo propuesto en este artículo mencionado (Sundell et al.) está directamente relacionado con la percepción y detección de microcalcificaciones y la evaluación de calidad de imagen, enmarcando dentro del paradigma de detección de señales a través de la capacidad de discriminación de ausencia o presencia de microcalcificaciones.(63)

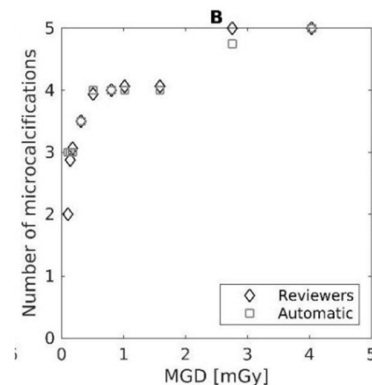


Fig 4. Cantidad de microcalcificaciones visible en función del MGD.

En la figura se visualiza la cantidad promedio de microcalcificaciones por ambos métodos a distintos niveles de dosis glandular. Esta muestra un incremento progresivo en la detección correspondiente a un aumento de la dosis, evidenciando que una mayor exposición mejora la relación señal-ruido. Se puede observar que el método automático detecta sistemáticamente una cantidad ligeramente superior de microcalcificaciones, esta diferencia sugiere que el algoritmo posee una mayor sensibilidad. Es importante notar, como verificación de lo comentado al final de la sección dedicada al método de Huda, que la curva obtenida tanto con el método manual basado en observadores como con el automático, la curva es bifásica (Fig. 4): se puede observar una meseta que no se corresponde teóricamente al aumento de visualización conforme aumenta la dosis glandular, esto se

podría explicar debido a limitaciones del método score de ACR para medir calidad de las imágenes.

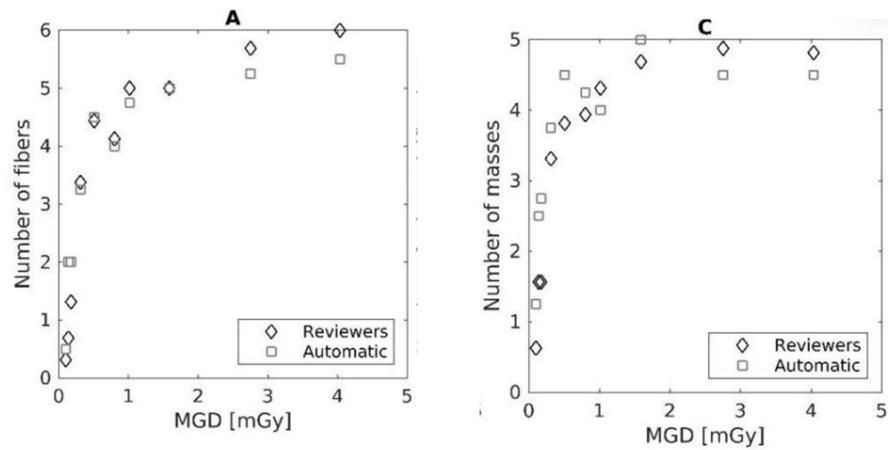


Fig. 5. Resultados obtenidos para fibras (Panel A. Izquierda) y masas (Panel B. derecha).

Otros resultados obtenidos para fibras y masas se grafican pudiendo observar (Fig. 5) que esta ya no es bifásica, mostrando un aumento de la visualización de estos objetos correlativo al aumento de la dosis que es lo esperado teóricamente, sin presentar una meseta como ocurre en la visualización de microcalcificaciones. Esta limitación del método ACR ya había sido antes detectada por Huda(2002).

Hipótesis

El mejor método basado en observadores humanos de medición de la calidad de las imágenes de microcalcificaciones en fantasmas mamográficos debe sustentarse de la identificación del píxel más intenso para cada micro visible excepto que el tamaño e intensidad de estas superen cierto valor que dificulte la búsqueda.

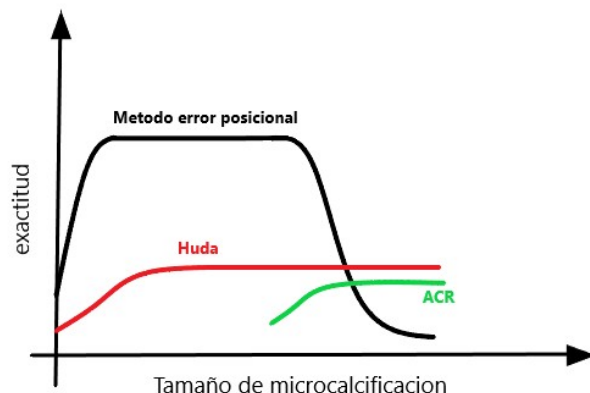


Fig. 6. Representación gráfica de hipótesis.

Basándonos en el principio formulado por Shannon(61), el método más exacto será el que contenga una mayor información de la señal original (fantoma). En este sentido el método de error posicional es el más completo en cuanto a información (a menos que se intentara además de detectar y localizar, reportar la intensidad observada de la señal comparada con una escala), por lo que decanta en que sea más exacto ya que contiene información de la localización de las microcalcificaciones. Los métodos tradicionales propuestos por Huda y ACR implican una reducción de información que se limita a la presencia o ausencia de estas microcalcificaciones lo que lleva a una pérdida de exactitud.

En la representación gráfica (*Fig. 6*) se observa que a medida que el tamaño de microcalcificaciones aumenta desde valores muy pequeños que permiten dependiendo del observador su visualización, la exactitud del método de error posicional aumenta hasta alcanzar un rango de desempeño estable y óptimo donde se pueden ver y localizar los píxeles más intensos. Una vez superado un tamaño la exactitud de este método disminuye considerablemente esto se debe a una mayor extensión de la señal lo que genera múltiples píxeles seleccionables de máxima intensidad lo que incrementa el error posicional reduciendo la exactitud del método. Por lo contrario en los métodos tradicionales mencionados con anterioridad, la exactitud se ve incrementada con el aumento del tamaño de las microcalcificaciones ya que la detección visual y el conteo se facilita, con estas señales tan grandes. Estos métodos superan en exactitud y rendimiento al método propuesto de error posicional. Esto lleva a identificar una limitación del método de autoría vinculada a la sensibilidad de variación espacial en señales de gran tamaño, lo que sugiere una combinación de métodos para mantener la exactitud en todo el rango de tamaños posibles.

Objetivos

Generar y poner a prueba nuevos métodos para evaluar la calidad de las imágenes de microcalcificaciones en mamografía digital 2D indirecta, utilizando fantasmas mamográficos ACR, que permitan mejorar la evaluación de los servicios imagenológicos.

Preguntas de investigación

Asumiendo que cada observador identificará distintos píxeles como los más intensos para el mismo set de imágenes:

- ¿Qué distribución asumirá la posición del píxel más intenso?
- ¿Qué tipos de errores pueden afectar el rendimiento de este método?
- ¿Es el “método de error posicional” el mejor método?

Diseño

Nuestra investigación es del tipo experimental, ya que presenta una estructura planificada que organiza cómo se desarrolla la misma, estableciendo de forma precisa qué condiciones se controlan. El propósito es asegurar que los datos obtenidos permiten comparaciones correctas y reducir la influencia de factores externos del estudio. Siguiendo a Cochran y Cox, este tipo de diseño implica “definir con exactitud lo que se hará para llevar a cabo el experimento”, acotando el concepto a los elementos que realmente permiten controlar la variabilidad y garantizar la validez de las conclusiones(65).

En el contexto de la evaluación de la calidad de imágenes como es el caso, este tipo de diseño permite determinar cómo las modificaciones en los estímulos causadas por ejemplo por variaciones en la técnica de adquisición, afectan los juicios de los observadores humanos, cada uno de ellos con sus desempeños característicos. El juego de imágenes que da origen a esta investigación fue adquirido por autores y colaboradores en distintos servicios y variadas condiciones de adquisición, como del receptor y de dosis. Este set de imágenes de fantoma ACR fue obtenido en tres servicios mamográficos del área metropolitana de nuestro país, para maximizar la variabilidad, se alternaron combinaciones de variables tanto modificables como de registro, se aplicaron dos modalidades técnicas una manual: (voltaje el tubo entre 24-32 kVp y corriente del tubo entre 80-140 mAs) y Semiautomática (voltaje el tubo fijo y corriente del tubo entre 35-235 mAs). Otros parámetros modificados experimentalmente fueron la limpieza del chasis (interna y externa), el tamaño de imagen y las horas de encendido del equipo, con el objetivo de generar imágenes de baja y alta dosis para el análisis(66).

A su vez es randomizado. Todos los observadores evaluaron el mismo set de imágenes pero en una secuencia aleatoria distinta para cada uno. Esta aleatorización individualizada busca evitar que el observador desarrolle patrones de respuesta o sea influido por expectativas previas, también busca independencia entre los observadores y disminuir sesgos como los que puede producir la fatiga visual.

Por último se caracteriza por presentar una asignación a ciegas (ciego). Los observadores no conocían las condiciones de adquisición de las imágenes a observar, esto es imprescindible para asegurar la objetividad en la investigación. Esta asignación se encuentra a cargo de autores que no participaron en la visualización de dichas imágenes. En síntesis, el diseño es experimental, randomizado y ciego, para fortalecer la validez científica del estudio, reduciendo el sesgo, mejorando la confiabilidad de los resultados y garantizando que las diferencias obtenidas no se atribuyen a factores subjetivos o contextuales.

Materiales y métodos

Como material indispensable dispusimos de un subconjunto de imágenes de fantasmas mamográficos del ACR tomado como muestra representativa del set original del trabajo de tesis de maestría de Maria Sarasua (n = 155 de 280). Las imágenes de fantasmas

mamográficos fueron visualizadas en el servicio de imagenología del Hospital de Clínicas Dr. Manuel Quintela. Se utilizó un monitor Barco Nio 5MP LED de 2560 x 2048 píxeles de resolución y 1200:1 de contraste, por lo que la visualización fue precisa y confiable, con herramientas tecnológicas avanzadas y en buen funcionamiento para un mejor manejo de detección por parte de los observadores(34)(33). Cada observador registró datos en tablas creadas en Hoja de Cálculo de Google individuales o en papel y luego se completó el registro en la nube.

A partir del estudio del marco teórico y de los antecedentes presentados (*ver Antecedentes, pág.19*), proponemos los métodos de medición de la calidad de imágenes de fantasmas mamográficos del ACR de abajo. Estos métodos, que son de nuestra autoría, fueron implementados en cascada a partir del primero. Esto quiere decir que el primero tiene suficiente información para implementar todos los otros y esto implica por definición, que son sistemas de medición de exactitud decreciente. Aunque los cuatro nuevos métodos fueron implementados en cascada vale la pena mencionar que fueron ideados en orden inverso (río arriba), partiendo del método de Huda y cols (2), así decidimos presentarlos.

Método 1. “N°. De Micros Visibles Esperadas y Detectadas Por Grupo”: En este método se propone cuantificar cuántas microcalcificaciones se ven en cada grupo por imagen, teniendo en cuenta la posición de cada una, ya que no se pueden interpretar independientemente del resto de los juicios. Comparando este nuevo método con el método de Huda obtenemos información adicional sobre el grupo al cual pertenecen las microcalcificaciones visualizadas. Por las propias características físicas del fantoma ACR y las teorías de percepción visual se espera que un grupo de microcalcificaciones con mayor cantidad de material, causará un estímulo de mayor intensidad y por tanto presentará un número de microcalcificaciones observadas mayor o al menos igual que grupos compuestos por objetos de menor cantidad de material e intensidad de estímulo (señal). Si ocurriera lo contrario, estaríamos ante observaciones imprevistas en la teoría, que prevé que un aumento de la señal genera un aumento en la relación señal ruido (SNR, ver ec. x y el apartado en SDT) mayor en la imagen o en el registro de los observadores. Este método nos lleva a dudar de (agrega incertidumbre a) una curva de N° de microcalcificaciones visibles en función de la dosis donde se incluyen observaciones anómalas.

El método descansa inicialmente en juicios independientes sobre resultados dicotómicos: visible o no visible. Para llevar a cabo una evaluación de la exactitud de este método nos podríamos respaldar en el paradigma de clasificación de Neyman-Pearson este permite decidir entre ausencia (H_0) o presencia de señal (H_1) controlando la probabilidad de falsos positivos. Estableciendo que para un nivel fijo de error de tipo 1(α) la regla de decisión más fuerte es la que maximiza la detección verdadera(68), este paradigma es muy utilizado en el ámbito de la detección debido a su capacidad para optimizar sensibilidad bajo un riesgo controlado.(48) Este paradigma se relaciona directamente con las curvas ROC (Receiver Operating Characteristic) esta representa gráficamente todas las posibles combinaciones entre la tasa de verdaderos positivos(sensibilidad) y la tasa de falsos positivos que se obtienen al variar el criterio o umbral de decisión de un sistema o observador, desde esta perspectiva de Neyman y Pearson se permite identificar para un valor de falsos positivos y el clasificador que ofrece la mayor sensibilidad posible.(48,69)

Método 2. “N°. Visible De Micros ESPERADA Por Grupo e Ítem” a diferencia del método anterior este nuevo método aporta información por microcalcificación individual, asignando una letra a cada una de ellas en su correspondiente grupo en orden alfabético y sentido horario. Este ítem adicionado a cada microcalcificación nos brinda un punto de referencia en la forma pentagonal establecida en el grupo. Esta información individual por microcalcificación nos permite comparar entre observadores cual fue la que visualizo cada uno ya que en el anterior método no se discrimina por microcalcificación individual observada sino el total

registrado por grupo. Se caracteriza en que cada microcalcificación del mismo grupo tiene la misma probabilidad de ser identificada, en el caso de un grupo donde la visualización se dificulta y los observadores presentan discrepancias en las microcalcificaciones observadas esta probabilidad se mantiene, estaríamos ante un problema si los observadores sistemáticamente observan las mismas microcalcificaciones ya que la probabilidad no sería la misma en todas ellas.

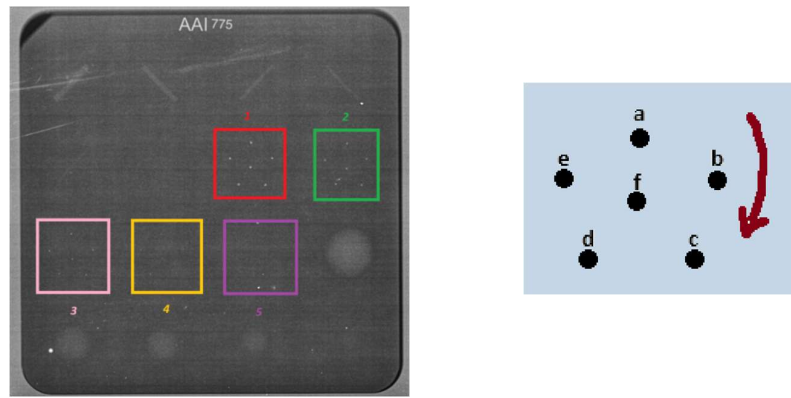


Fig. 7. Panel A.(Izquierda) grupos marcados del 1 al 5. **Panel B. (derecha)** representación de un grupo de microcalcificaciones con su correspondiente ítems en cada una de ellas.

Método 3. “N°. Visible De Micros Por Coordenadas” siguiendo con la metodología utilizada en el método “N° Visible De Micros ESPERADA Por Grupo e Ítem” donde se asigna un ítem a cada microcalcificación, el observador localizara en el espacio el píxel de mayor intensidad de cada una de ellas obteniendo coordenadas en los dos ejes x e y. Este nuevo método aporta información de localización en el espacio de cada microcalcificación en su grupo correspondiente. A través de este método se logra determinar todas las categorías correspondientes al paradigma de Neyman-Pearson mediante el método de decisión por mayoría, los falsos positivos serán aquellos registros que luego de obtenido su valor de desvíos estándar calculado a partir de las coordenadas registradas por cada observador, este se aleje tres desvíos estándar de la media para ese registro de todos los observadores, los falsos negativos se van a distinguir ya que no tendremos coordenadas en el registro de un observador pero la mayoría de los observadores si presentan registro para ese caso. Un verdadero positivo se determinarán ya que la mayoría de los observadores visualizan y el registro no se aleja tres desvío estándar de la media de todos los observadores y por último los verdaderos negativos se determinan si la mayoría de los observadores no observan un caso.

En conjunto los tres métodos desarrollados están sustentados por el paradigma de Neyman-Pearson, donde todos buscan cuantificar y clasificar la respuesta de los observadores en cuanto a la detección correcta o errónea, permitiendo la distinción entre falsos positivos y falsos negativos, verdaderos negativos y falsos negativos. Con este paradigma se pueden interpretar los resultados obtenidos desde una perspectiva estadística, donde la toma de decisiones por cada observador es evaluada en función de su probabilidad de corresponder a una el real o a una respuesta creada por azar. Así mismo, los métodos propuestos no solo garantizan medir la visualización de las microcalcificaciones, sino que consolida cuantitativo donde que permite analizar la precisión perceptual de los observadores.

Método 4. “Método De Error Posicional” para los métodos anteriormente mencionados, se utilizan datos discretos y dicotómicos, ya que es ver o no ver microcalcificaciones, este nuevo y último método introduce un enfoque cuantitativo y continuo centrado en la precisión espacial de la percepción visual de los observadores. En este no solo se analiza si la

microcalcificación fue detectada sino que tan bien se identifica la posición de esta en la imagen. La información de posición para este método se obtiene también del método “N° Visible De Micros Por Coordenadas” esto permite construir un modelo continuo del error posicional considerando la distancia entre las posiciones registrada por cada observador y la posición de referencia real en este caso determinada por el cálculo del promedio de las coordenadas de todos los observadores, por lo que el error posicional se expresa mediante una medida continua que cuantifica la desviación espacial en unidades de pixel. Este análisis ofrece nueva información sobre la precisión de los observadores y la calidad de la imagen, y junto a la dosis utilizada en las adquisiciones se pueden establecer modelos predictivos del error posicional esperado. Este método más complejo nos aporta nueva información de detección y localización de microcalcificaciones en fantasmas mamográficos ACR

Se completará manualmente la tabla correspondiente al método “N° Visible De Micros Por Coordenadas” y luego a partir de esta se completará automáticamente a través de fórmulas las planillas correspondientes a los métodos restantes.

Resultados

Este trabajo se centra en la presentación de nuevos métodos y no pretende abordar de manera definitiva todas las problemáticas involucradas en estos, sino que su propósito es ofrecer un primer acercamiento exploratorio que permita reconocer posibles ventajas y aspectos de estas propuestas metodológicas. Para facilitar su comprensión, los resultados se estructuran en diez secciones.

Evaluación de la posible utilidad del método 1. Para poder realizar este método de acuerdo con el paradigma de Neyman y Person es necesario determinar los verdaderos y falsos positivos y negativos (VP, FP, VN y FN). La definición es frecuentemente posible por el criterio de la mayoría (del consenso mayoritario), ya que disponemos de hasta 5 juicios independientes (5 observadores). A partir de esto podemos construir curvas ROC para cada imagen del set. Estas curvas son de dos tipos: las crudas y las corregidas según SDT. Decidimos poner a prueba el método con datos simulados. Tómese esta situación como ejemplo posible:

Tabla I. Observaciones hipotéticas/simuladas.

Imagen i	juicio de cada observador (número de micros visibles)				
Grupo de Micros	O1	O2	O3	O4	O5
1	6	6	6	6	6
2	5	5	4	5	6
3	6	6	6	6	5
4	4	4	2	3	4
5	2	1	0	1	2

Si consideramos la Teoría de la Detección de Señales nos daremos cuenta que el grupo 3 de la tabla anterior es imposible (no esperamos mayor detectabilidad con menor estímulo).

Definimos k como el número de micros visible por grupo según cada observador y también definimos el consenso mayoritario que llamamos “ m ” como el estándar de oro perceptual (en inglés “Gold Standard”, GS) usando la regla de la mayoría para cada grupo:

Tabla II. Consenso mayoritario m .

Grupo de Micros	m
1	6
2	5
3	6
4	4
5	No se puede decidir sobre m

Si el consenso mayoritario es 5 (como en el grupo 2), esto quiere decir que se debieron detectar 5 microcalcificaciones. Si k para el observador 1 fuera 6, entonces en ese caso el O1 detectó esperadamente 5 y detectó 1 más por error, y a esta la consideramos como FP. Asumimos entonces que m es lo esperado y a partir de ello se definen desviaciones. En otras palabras:

$m - k$ es el número de FN, si da negativo (es decir k es mayor que m), entonces son FP no FN

$$VP = \min(k, m)$$

$$VN = 6 - VP - FN - FP$$

Ejemplos para un observador

$$m = 4, k = 6$$

$$FN = 4 - 6, \text{ no hay FN, hay 2 FP}$$

$$VP = \min(4, 6) = 4$$

$$VN = 6 - 4 - 0 - 2 = 0$$

Ejemplos para otro observador

$$m = 4, k = 3$$

$$FN = 4 - 3, 1 \text{ FN, no hay FP}$$

$$VP = \min(4, 3) = 3$$

$$VN = 6 - 3 - 1 - 0 = 2$$

Esto permite obtener matrices de confusión para cada grupo de microcalcificaciones, siempre y cuando se pueda calcular m (lo que no siempre es posible con esta regla del consenso mayoritario).

Tabla III. Matriz de confusión para el grupo 1.

—	Se ve	No se vé	Total
Se debe ver	$6 \times 5 = 30$ micros (VP)	0 (FN)	30 (100%)
No se debe ver	0 (FP)	0 (VN)	0 (0%)
Total	30 (100%)	0 (0%)	—

Otro ejemplo de acuerdo a la *Tabla I*:

Tabla IV. Matriz de confusión para el grupo 2.

—	Se ve	No se vé	
Se debe ver	$5 \times 3 + 4 + 5 = 24$ (VP)	1 (FN)	25 (83.33...%)
No se debe ver	1 (FP)	4 (VN)	5 (16.66...%)
	25 (83.33...%)	5 (16.66...%)	—

De acuerdo con los 150 juicios que se consideran en la *Tabla I*, se puede primero obtener la lista de juicios con su estándar de oro y la interpretación en términos de aportes a la matriz de confusión emergente, para finalmente, y luego obtener la curva ROC hipotética de la Fig. y:

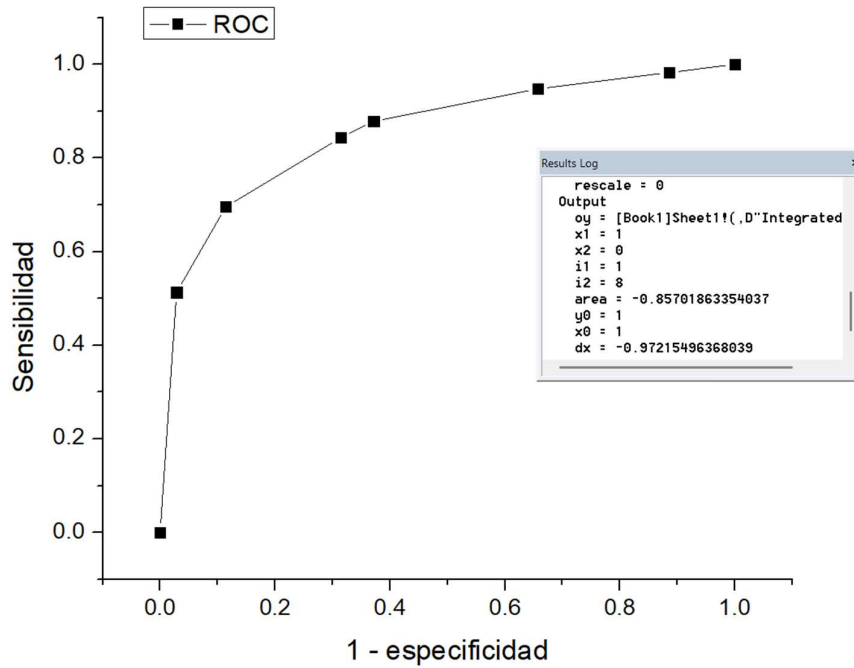


Figura 8. Curva ROC para el método 1 en condiciones de juicios independientes a partir de datos simulados/hipotéticos. Para esta curva ROC se usaron todos los datos de la Tabla I. El área bajo la curva ROC (AUC, del inglés Area Under the Curve) es 0.857

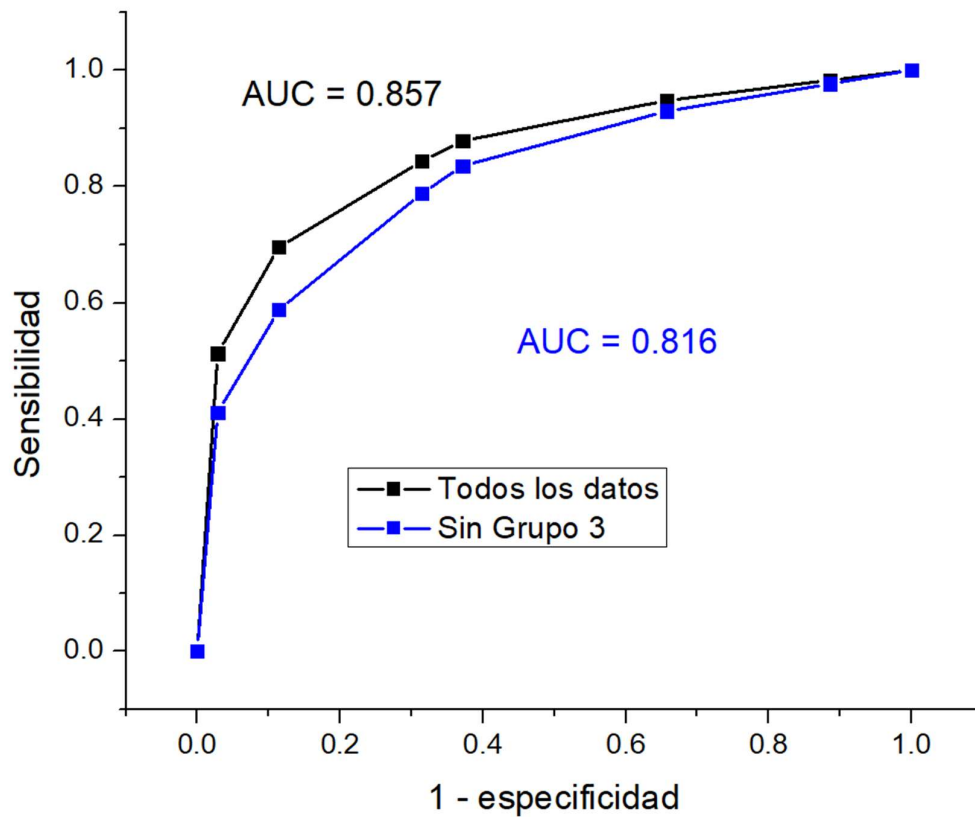


Figura 9. Curva ROC para el método 1 en condiciones de juicios independientes a partir de datos simulados/hipotéticos (curva negra) y eliminando los datos del grupo 3 (curva azul). La curva ROC en negro ya fue presentada en la Figura X. Aquí en azul se muestra la curva ROC resultante del análisis de los datos de la Tabla I pero sin considerar los datos del G3, ya que por teoría no esperamos mejores scores para el G3 que para el G2. El área bajo la curva ROC (AUC, del inglés Area Under the Curve) es 0.857 para la curva negra y 0.816 para la azul.

Lo que proponemos a través del Método 1, es que el uso de curvas ROC consistentes con SDT, es decir curvas que se elaboran luego de la eliminación de datos que contradicen las leyes de la detección de señales, son mejores que las curvas ROC crudas como forma de medir la calidad de las imágenes. De lo contrario, es esperable una sobreestimación, un sesgo positivo en los resultados de los métodos basados en observadores para el ejemplo hipotético elegido. Aquí se muestra en un caso simulado, que la eliminación de datos del grupo 3 condujo a una reducción del AUC de $0.857 - 0.816 = 0.041$ (es decir de un 5%).

Características del proceso de registro realizado y del conjunto de datos obtenidos.

Cada observador registró en Hoja de Cálculo de Google el posicionamiento de cada microcalcificación de cada grupo en el eje "X" y en el eje "Y" en píxeles. La matriz original de datos es de 6x5x5x155 registros: 6 micros por cada uno de los 5 grupos (30 microcalcificaciones por imagen) a evaluar por 5 observadores en 155 imágenes, un total de 23250 observaciones, las que generaron dos tipos de registros: NV (No visto) o coordenadas x e y en píxeles, totalizando 46500 registros. (ver Anexo 1)

Análisis de datos: medidas de centralidad y de variabilidad de la posición. La medida de centralidad utilizada en este caso es la media que se define como el cociente entre la suma de todos los valores de una variable y el número total de observaciones. Este valor representa el punto central de la distribución, aunque presenta como limitación su sensibilidad ante la presencia de valores atípicos, los cuales pueden modificar significativamente su valor. Por otra parte, la desviación estándar cuantifica el grado de variabilidad de los datos respecto a la media, se define como la raíz cuadrada positiva de la varianza y permite interpretar cuánto se alejan, en promedio, los valores observados del promedio general. Valores bajos de desviación estándar indican una baja variabilidad, es decir, que los datos están concentrados alrededor de la media, mientras que valores altos reflejan una mayor dispersión. La elección de estas medidas se fundamenta en la suposición de que la distribución de los valores es aproximadamente normal, lo cual se respalda en trabajos previos consultados donde también se medía la posición en píxeles y se aplicaban estas bajo el mismo supuesto de normalidad.

Los resultados que siguen se centran en la investigación exploratoria del Método 4, correspondiente a la medición del error posicional como indicador perceptual de la calidad de las imágenes. Estudiamos la precisión y la veracidad con que los observadores localizan la posición de las microcalcificaciones a través de la identificación de un único píxel para cada microcalcificación en las imágenes. Como se esperaba, al analizar los registros obtenidos se observó que las posiciones marcadas por los observadores presentan variabilidad (Fig 10. Panel A, (izquierda) Panel B (derecha) y Tabla V). Cada uno de estos definió las coordenadas en píxeles, mostrando coincidencias en algunos casos y discrepancias en otros.

Para el análisis de estos datos, se calcularon la media y la desviación estándar como medidas de centralidad y variabilidad respectivamente (Tabla V).

Tabla V. Registro

	imagen_047.tif Grupo 1											
	micro a		micro b		micro c		micro d		micro e		micro f	
	Xa	Ya	Xb	Yb	Xc	Yc	Xd	Yd	Xe	Ye	Xf	Yf
media	3308	3166,6	3397,8	3224,8	3366	3327,6	3262,2	3328,8	3228	3231,2	3312,4	3265,8
DE	1.2247	1.673	1.789	0.837	1.225	1.140	0.447	4.970*	0.707	0.836	2.074	1.095

* outlier

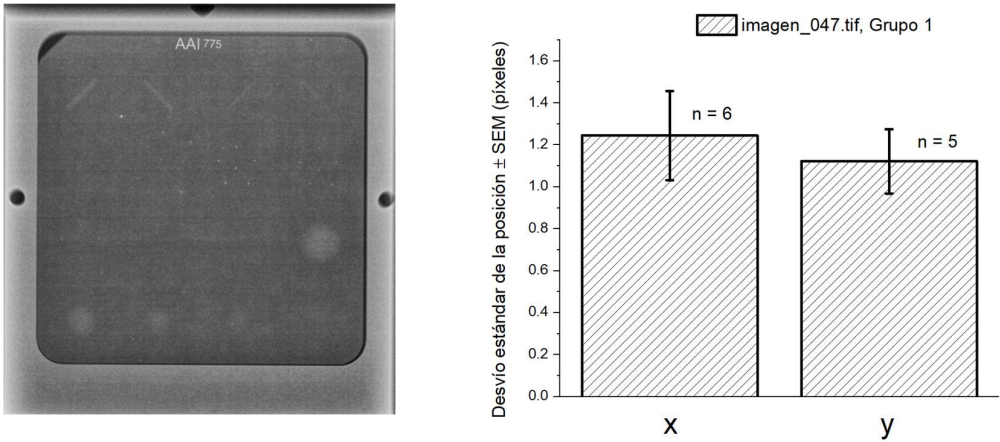


Figura 10. Ejemplo de imagen y de datos de localización de las microcalcificaciones del grupo 1. Panel A. (izquierda) Se muestra una imagen representativa de aquellas de buena calidad (MGD = 2.05 mGy). Se puede apreciar que las microcalcificaciones simuladas del grupo 5 son prácticamente indistinguibles del fondo. Las microcalcificaciones del grupo uno son notoriamente visibles (todos los observadores reportan sus localizaciones). **Panel B. (derecha)** Gráfico de barras de los desvíos posicionales con sus correspondientes errores

estándar de la media (SEM). A pesar de que todas las microcalcificaciones simuladas del grupo 1 son visibles en esta imagen del fantoma, la tarea de localización de las mismas se realiza con error.

Evaluación de la distribución de la variabilidad posicional y separación de errores. A través de los datos obtenidos anteriormente se logró graficar dos histogramas en los que se observa, en el eje Y, la frecuencia absoluta de los datos, mientras que en el eje X se representan los valores de desviación estándar. Los datos utilizados en cada histograma corresponden a los desvíos estándar calculados para cada grupo y para cada eje de coordenadas, obteniéndose así dos gráficos por grupo. El tamaño muestral total fue de 930, conformado por seis valores (uno por cada microcalcificación por grupo) en un eje determinado, multiplicado por 155 correspondiente al total de imágenes que componen el set de datos.

En el histograma con escala lineal (Fig. 11) sin aplicar transformación logarítmica, no fue posible visualizar con claridad la distribución de la variabilidad posicional. Esto debido a la existencia de valores extremadamente bajos y valores muy elevados, que distorsionan la representación. Para resolver esta limitación, se aplicó una transformación logarítmica en base 10 (Fig. 12) a los datos obtenidos de variabilidad, lo que permitió mejorar la visualización de la distribución.

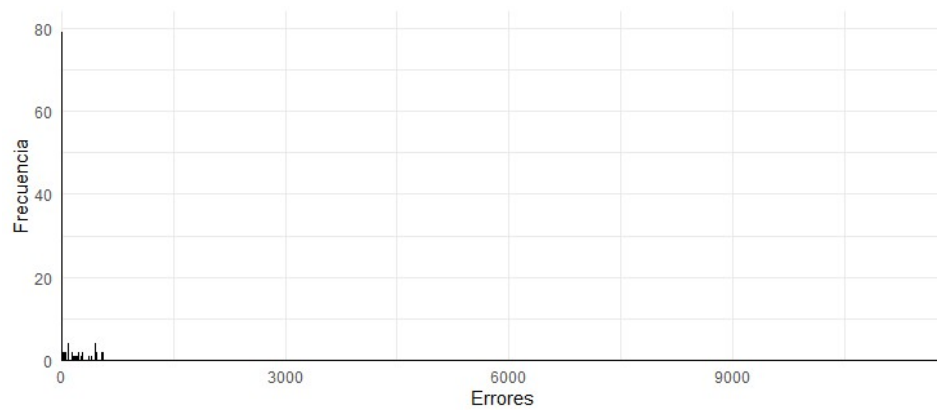


Figura 11. Histograma desvío estándar con escala lineal (sin transformación logarítmica).

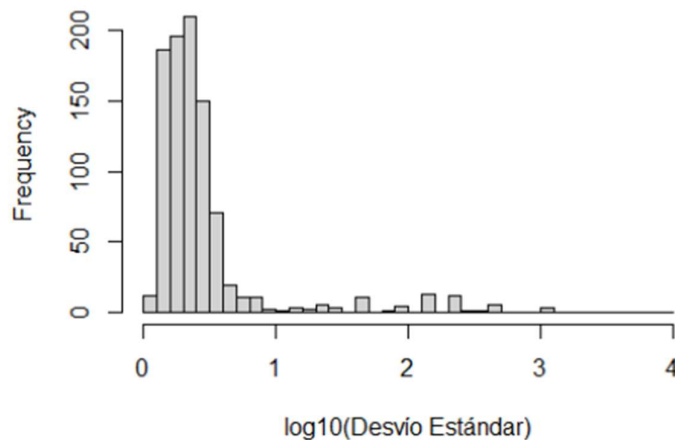


Figura 12. Histograma desvío estándar con transformación logarítmica.

Al observar los histogramas con valores en $\log_{10}$, se aprecia que para valores bajos de desviación estándar la distribución toma una forma aproximadamente normal. Sin embargo, hacia la derecha se extiende una cola larga, lo cual indica la presencia de valores con alta variabilidad. Dado que estos valores no podían atribuirse a una única fuente de error (errores posicionales), se interpretó que existía otra fuente de error, donde se llegó a la conclusión que eran errores de registro. El análisis detallado de los datos de los cinco observadores permitió identificar que, en determinados casos, uno de ellos registraba una coordenada muy diferente y alejada de las restantes, concluyendo que estas discrepancias se deban a errores humanos en el proceso de registro, al registrar en la celda correspondiente en la Hoja de Cálculo de Google valor de coordenada equivocado.

Adicionalmente se observó que cuando el procedimiento de registro se realizaba en dos etapas, primero anotando las coordenadas en papel y luego transcribiéndolas a la Hoja de Cálculo de Google, los errores aumentaban significativamente. Estos resultados permiten dar respuesta parcial a la pregunta de investigación ¿Qué tipos de errores pueden afectar el rendimiento de este método? En este caso, errores de registro fueron identificados como fuentes relevantes de variabilidad. Para avanzar en la respuesta a la pregunta de investigación, ¿Qué distribución asumirá la posición del píxel más intenso? Es necesario depurar los datos eliminando los valores considerados como errores de registro. Con este fin, se procedió a ajustar modelos a los histogramas de los valores logarítmicos de la desviación estándar, con el objetivo de establecer un criterio estadístico que permitiera distinguir entre los valores de errores posicionales y aquellos correspondientes a errores de registro.

Nuestra hipótesis de trabajo propone que la frecuencia absoluta de los valores de $\log_{10}$ de la desviación estándar se puede modelar como la suma de una distribución normal y una recta horizontal, más el error estándar. Este modelo, el más simple con cuatro parámetros, asume que la distribución normal describe los valores bajos del $\log_{10}$, correspondientes a errores posicionales asociados al proceso de localización de microcalcificaciones mientras que la recta horizontal con pendiente igual a cero describe los valores altos del $\log_{10}$, asociados a errores de registro.

La hipótesis podría expresarse como:

$$y = \text{Normal} + \text{Recta (pendiente} = 0) + \text{error estándar}$$

Modelo que representa la hipótesis ajustada. Como enunciamos en la hipótesis presentada anteriormente en el histograma presentado (Fig. 13) se observa que el modelo propuesto presenta un ajuste razonable a la distribución de densidad de probabilidad empírica derivada de los datos. Esta correspondencia sugiere que el modelo de distribución normal más una recta horizontal logra describir de manera adecuada el comportamiento general de los valores del $\log_{10}$ de los desvíos estándar, tanto la región asociada a los errores posicionales como la cola extendida vinculada a los errores de registro.

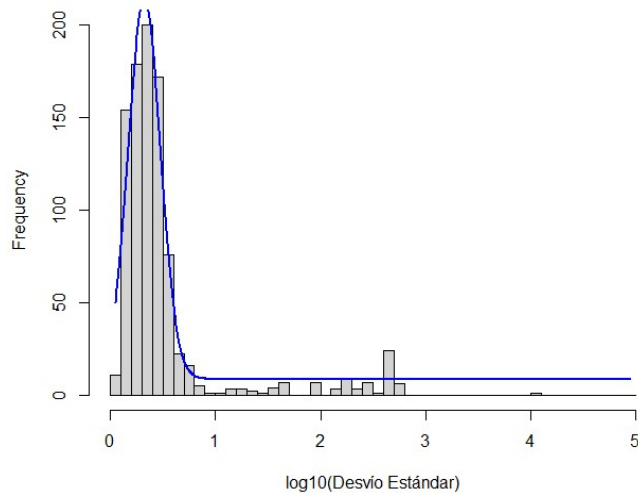


Figura 13. Histograma desvío estándar con transformación logarítmica y modelo ajustado que representa la hipótesis.

Variación intergrupar y calidad del ajuste del modelo estadístico. Al analizar los histogramas correspondientes a cada grupo (Fig.14), se observa que en el Grupo 1 (Panel A) el pico de frecuencia es el más alto, superando los 200. En el Grupo 2 (Panel B) Se evidencia una disminución en la altura del pico, aunque este sigue siendo mayor que el del Grupo 3 (Panel C), el cual, a su vez, presenta una frecuencia más alta que la del Grupo 4 (Panel D). Esta tendencia se explica por la diferencia en la cantidad de observaciones registradas por grupo, siendo el Grupo 1 (Panel A) el que cuenta con el mayor número de datos. El Grupo 5 no fue considerado para el análisis debido a que no disponía de un número suficiente de observaciones, a consecuencia del tamaño reducido de las microcalcificaciones, lo cual no permitió realizar un ajuste estadístico.

Las diferencias en el tamaño muestral se refleja directamente en la calidad del ajuste del modelo. Se observó que el ajuste del modelo propuesto (distribución normal + recta horizontal) es razonablemente bueno en los Grupos 1, 2 y 3, mientras que en el Grupo 4 el ajuste resulta menos satisfactorio, mostrando desviaciones más notorias entre la distribución teórica y la empírica. Este comportamiento podría mejorarse mediante el empleo de un modelo alternativo capaz de representar adecuadamente la mayor dispersión de los datos observados en dicho grupo.

La segunda parte de la hipótesis plantea que los errores de registro son constantes en todos los grupos, ya que no depende de la calidad visual, sino del hecho de registrar las coordenadas obtenidas. Por lo que la recta horizontal del modelo (con pendiente cero) debe mantenerse constante entre los grupos. En cambio, la distribución normal asociada a los errores posicionales sí debe ajustarse individualmente para cada grupo, ya que su forma y dispersión dependen directamente de la calidad de las imágenes y de la localización de las microcalcificaciones.

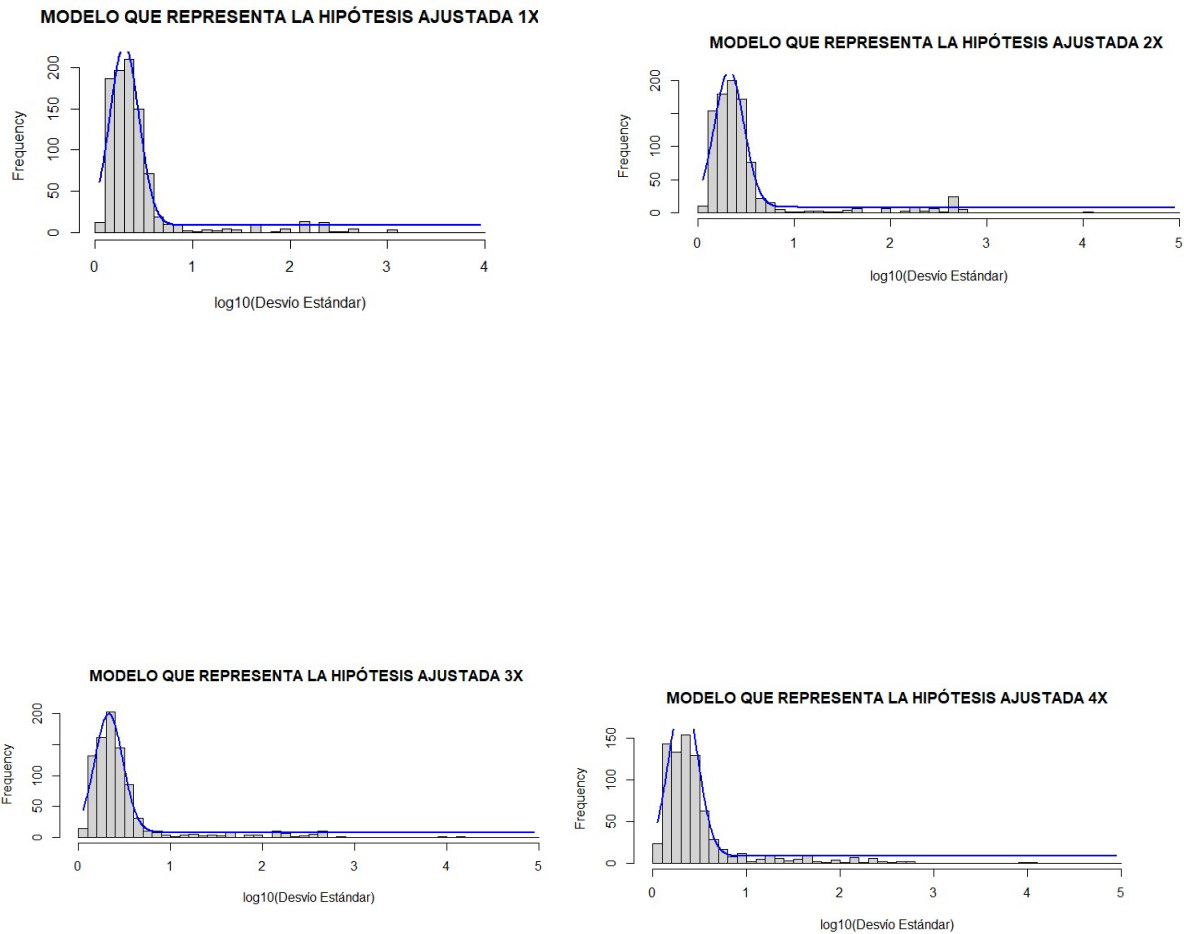


Fig. 14. Histograma desvío estándar con transformación logarítmica y modelo ajustado que representa la hipótesis. Panel A. (superior-izquierda), Panel B. (superior derecha), Panel C. (inferior-izquierda), Panel D. (inferior-derecha)

Interpretación de los parámetros del modelo lineal horizontal. En la tabla (Tabla VI) presentada se detallan los parámetros del modelo lineal horizontal, intercepto, error estándar y n. Correspondientes a todos los grupos y a sus respectivos ejes de coordenadas. A partir de estos valores se obtuvo posteriormente el promedio conjunto de los parámetros en los ejes X e Y. El intercepto representa la media muestral de las observaciones, ya que el modelo ajusta una recta horizontal situada a la altura del valor medio de los datos. Dicha recta posee pendiente cero, ya que nuestra hipótesis planteada es que la probabilidad de cometer un error de registro se mantiene constante a lo largo de todo el rango observado, independientemente al grupo considerado. Por lo tanto, este componente del modelo refleja un nivel uniforme de error que se asume constante entre los grupos 1 a 4. El error estándar del intercepto cuantifica la precisión de la estimación de dicho parámetro. Valores bajos de este indicador implican que la media está determinada con alta exactitud y que la dispersión de las observaciones alrededor de la recta horizontal es baja. Un error estándar bajo respalda la estabilidad del modelo y refuerza la validez de la hipótesis que sostiene que es constante el error de registro entre los distintos grupos.

Tabla VI. Parámetro del modelo lineal horizontal

	Grupo 1		Grupo 2		Grupo 3		Grupo 4		Grupo 5		Promedio		Promedio
	x	y	x	y	x	y	x	y	x	y	x	y	x e y
Intercepto	5,82	6,00	11,71	10,00	11,29	9,43	10,00	4,36	1,50	0,50	9,71	7,45	8,58
Error	1,27	1,33	4,65	3,25	3,79	3,78	4,57	1,01	0,50	0,50	3,57	2,34	2,96
n	11,00	9,00	7,00	6,00	7,00	7,00	7,00	11,00	2,00	2,00			

Estimación del valor crítico (SD) e identificación de errores de registro. Se utilizó el programa R Studio para marcar el límite de los desvíos estándar crítico para así obtener un punto de corte óptimo por grupo de microcalcificaciones (*Figs.15*), con estos datos se detectaron los errores de registro de cada observador. La línea punteada horizontal en color rojo, con un valor de 11,54, está representada por el valor medio del intercepto (8,58), más un error estándar medio de (2,96). La adición de un error estándar se adoptó como criterio de compromiso metodológico, con el objetivo de minimizar la exclusión de datos que podrían corresponder a errores de posicionamiento, y no necesariamente a errores de registro. La intersección entre la curva normal ajustada y la línea horizontal define el valor crítico de desviación estándar (SD) utilizado como umbral de referencia. Los valores situados a la derecha de dicho punto se consideran estadísticamente improbables de atribuir a la variabilidad interobservador, por lo que se interpretan como posibles errores de registro. Estos casos fueron identificados y marcados para su exclusión en el conjunto de datos original, garantizando así una depuración objetiva del error posicional. A continuación mostramos gráficamente los resultados. (*ver anexo 2*)

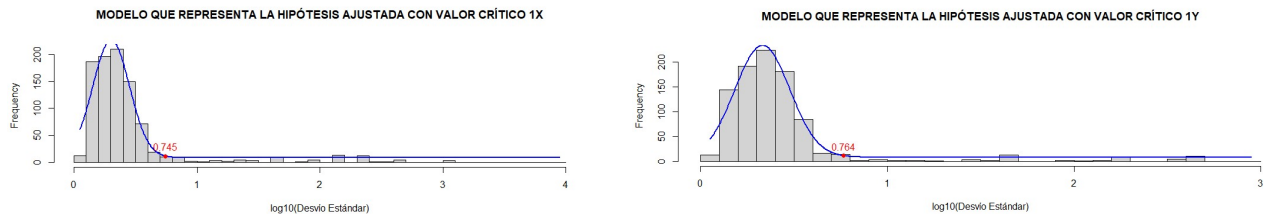


Fig 15. Histograma desvío estándar con transformación logarítmica y modelo ajustado que representa la hipótesis con valor crítico. Panel A. (izquierda) correspondiente a grupo 1x, Panel B. (derecha) correspondiente a grupo 1y.

Cálculo de valores críticos en Log10 y en pixeles del desvío estándar por grupo y eje.

En esta tabla (*Tabla VII*) se muestra el valor log base 10 del desvío estándar para cada grupo en su respectivo eje x e y, y luego los valores con la función inversa. Los valores de desvío estándar mayores al valor correspondiente para cada grupo y eje se marcan en los datos originales de cada observador y se consideran posibles errores de registro.(*ver Anexo 3*)

Tabla VII. Valores críticos en Log10 y en pixeles del desvío estándar

Grupo.	Eje de coordenadas.	Log10 Desvío estándar Crítico	Desvío Estándar Crítico (píxeles).
1	X	0,745	5,55
1	Y	0,764	5,81
2	X	0,771	5,90
2	Y	0,768	5,86
3	X	0,776	5,97
3	Y	0,781	6,04
4	X	0,781	6,04
4	Y	0,796	6,25

Error posicional en X vs en Y. En la Figura (Fig. 16) se comparan el logaritmo del valor del error posicional (en x y en y) en función de los cuatro grupos de microcalcificaciones, cada uno con un nivel distinto de visibilidad y cantidad de material radiopaco. Ambos conjuntos de datos se ajustaron mediante un modelo exponencial, observándose un buen nivel de ajuste (R^2 ajustado $\approx 0,98$), lo que muestra que ambas curvas responden a la misma familia de funciones. La curva negra (x) alcanza un máximo cercano a 0,78, indicando un incremento del error posicional a medida que aumenta el número de grupo, posiblemente asociado a una limitación en la calidad de imagen, esto sugiere que aunque las microcalcificaciones sean más visibles, el incremento en la densidad de material o en la intensidad del contraste puede generar zonas de saturación que dificultan la localización precisa del pixel central de máxima intensidad de la microcalcificación. Por su parte, la

curva roja (y) presenta diferencias, ya que comienza en valores iniciales más bajos, alcanzando niveles óptimos y estabilizándose, este refleja una limitación perceptual del observador, es decir una restricción en el sistema visual, más que por la calidad de la imagen, por lo que el aumento de material o intensidad no mejora la detección ni la precisión posicional, manteniéndose un valor crítico constante.

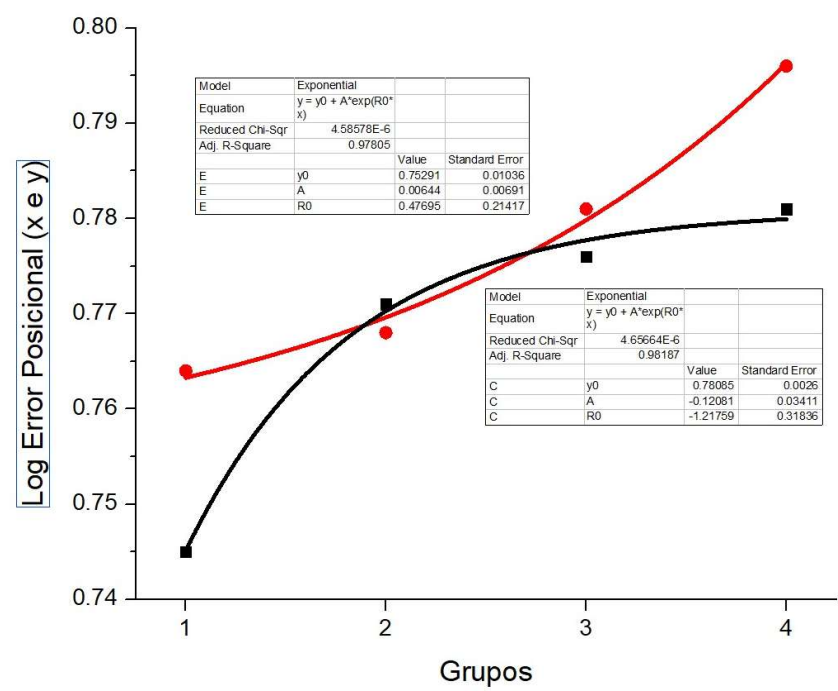


Fig. 16. Gráfica Log10 Error posicional en X vs en Y.

Discusión

En esta sección se integran los principales hallazgos del estudio, interpretando los resultados ante teorías perceptuales y métodos previos de evaluación de calidad de imagen. Asimismo, se analizan las limitaciones de la investigación y se proponen líneas de mejora para estudios futuros.

Un soporte fundamental para el enfoque tomado en esta investigación fue literatura previa donde tendía hacia métodos de evaluación que no se limitaban a características físicas, sino que integraba la percepción humana, estos proponen métodos alternativos y complementarios para la evaluación de la calidad de imagen en fantasmas mamográficos ACR, muchos de ellos basados en principios de la visión humana. Se han desarrollado estudios psicofísicos que analizan la respuesta de observadores humanos ante los objetos del fantoma, estudiando la variabilidad interobservador. A su vez, surgieron métodos automatizados que cuantifican objetos de la imagen y los correlacionan con el desempeño humano, buscando predecir la visibilidad de fibras, masas y microcalcificaciones en condiciones clínicas simuladas. Incluso, en los últimos años se han introducido sistemas basados en aprendizaje profundo, capaces de generar puntajes comparables a los de los observadores humanos a partir de imágenes del fantoma(70–72).

Uno de los primeros inconvenientes manifestados fue la alta variabilidad entre los registros de los observadores, lo cual no se lograba explicar por variabilidad de posición de las microcalcificaciones observadas por los distintos observadores. Para ello fue necesario definir cuál otra causa podría dar respuesta al aumento significativo de este valor, donde luego de investigar antecedentes en registros múltiples de observadores humanos, pudimos confirmar que esta alta variabilidad se debía a errores de los observadores al momento del registro del número de coordenadas obtenido (errores de registro). Para ello fue necesario definir un punto crítico cuantitativamente mediante script para el software Rstudio (ver Anexo 4) en donde poder separar este tipo de error de los errores posicionales, para ellos se ajustó la distribución normal + recta horizontal a los datos empíricos para poder definir este punto, los resultados obtenidos del modelo ajustado (distribución normal + recta horizontal) mostró una concordancia con la hipótesis planteada, los errores posicionales se distribuyen de manera normal demostrando la variabilidad interobservador en las tareas de percepción, y la recta horizontal manifiesta una tasa constante de errores de registro en los diferentes grupos. Esta concordancia de la hipótesis fue mayor en los grupos de microcalcificaciones de mayor tamaño e intensidad (grupos del 1 al 3) y menos estables en grupos de menor tamaño (grupo 4), representando que el nivel de detección decrece con la disminución del tamaño y la intensidad del estímulo, coincidiendo con teorías previas Carrasco 2001 (40) y Krupinski 1996 (58)) donde se destaca la relación entre la dificultad de la visualización y la precisión espacial del observador.

La dispersión observada en los registros demuestra que el desempeño perceptivo no depende solo de la calidad técnica de la imagen, sino de factores cognitivos. Por lo que los métodos de autoría tratan de integrar la calidad técnica de las imágenes y factores perceptivos de los observadores formando modelos más integrales de evaluación, incorporando modelos psicofísicos y estadísticos. Los métodos tradicionales como ACR y autores como Huda(2) se centran en la detección visual discreta (ver/no ver) de objetos simulados, estos métodos son relativamente simples pero infravaloran la complejidad de la

relación antes mencionada entre las características físicas y la percepción humana. El método de error posicional expresa la respuesta visual en términos de desviación continua, lo que constituyendo un avance metodológico comparado con escalas discretas, ya que la medida se interpreta en unidades de píxeles, lo que permite distinguir entre errores del sistema de adquisición y errores por limitaciones humanas, ya sea de registro o posicionamiento. Su aplicación podría mejorar la sensibilidad y especificidad de programas de control de calidad lo cual permitiría detectar problema sutiles en rendimiento de equipos y/o en condiciones de observación que métodos tradicionales no evidenciaron.

La primer sección de los resultados se da producto de datos simulados/hipotéticos ya que no son datos obtenidos por el registro auténtico de los observadores sino que permiten explorar la idea en principio de cómo evaluar el consenso por mayoría y la construcción de curvas ROC consistentes con los principios de SDT es conceptualmente correcto, estos datos no representan las condiciones reales de lectura sino que es probar la idea y visualizar cómo se modifican los índices ante las correcciones de inconsistencias con SDT, en los resultados hipotéticos se muestra que la eliminación de datos incoherentes (grupo 3) con los principios básicos de detectabilidad conduce a diferencias observables en las curvas ROC. Sin embargo al aplicar este método sobre datos reales será necesario tomar decisiones con mayor juicio, ya que la clasificación de VP, FP, FN y VN construida por consenso por mayoría podría volverse inestable para datos con alta variabilidad por lo que se deben establecer bandas de variabilidad y criterios de confianza para establecer contradicciones con las teorías psicofísicas.

Observando la Gráfica Log10 Error posicional en X vs en Y (Fig. 16) donde se visualiza una diferencia notable entre los valores errores posicionales en escala log10 en los ejes x e y, sigue lo propuesto por autores como Curcio y cols(39). Quienes demostraron que la distribución de los fotorreceptores en la retina no es homogénea sino que presenta una organización espacial diferencial entre el eje horizontal y vertical. Bajo este marco teórico sería esperable que los errores en x fueran menores que en el eje y, en coherencia con esta afirmación, la línea correspondiente a x (línea negra) muestra una tendencia a estabilizarse alcanzando una meseta en un valor de error posicional menor, por otro lado los errores posicionales en el eje y (línea roja) presentan valores superiores y se tiende a alcanzar la meseta a niveles más alto de error posicional, esto se correlaciona con una menor densidad de fotorreceptores en este eje, como se puede observar con la continuidad de la línea roja donde esperaríamos que se observe esta meseta pero a mayor valor de error posicional que en el eje x. También podemos observar que si se extrapolan los valores del eje x hacia niveles más bajos de error su forma se asemejaría a la curva del eje y, reflejando la relación esperada entre ambas funciones perceptuales. En conjunto, esta figura (Fig. 17) no solo describe el comportamiento de los errores posicionales en el presente estudio sino que también constituye una aplicación de principios de la organización retiniana. Esta comparación entre ambas curvas sugiere que las diferencias descritas por Curcio et al. se manifiestan directamente en tareas de localización espacial de microcalcificaciones en imágenes del fantoma ACR.

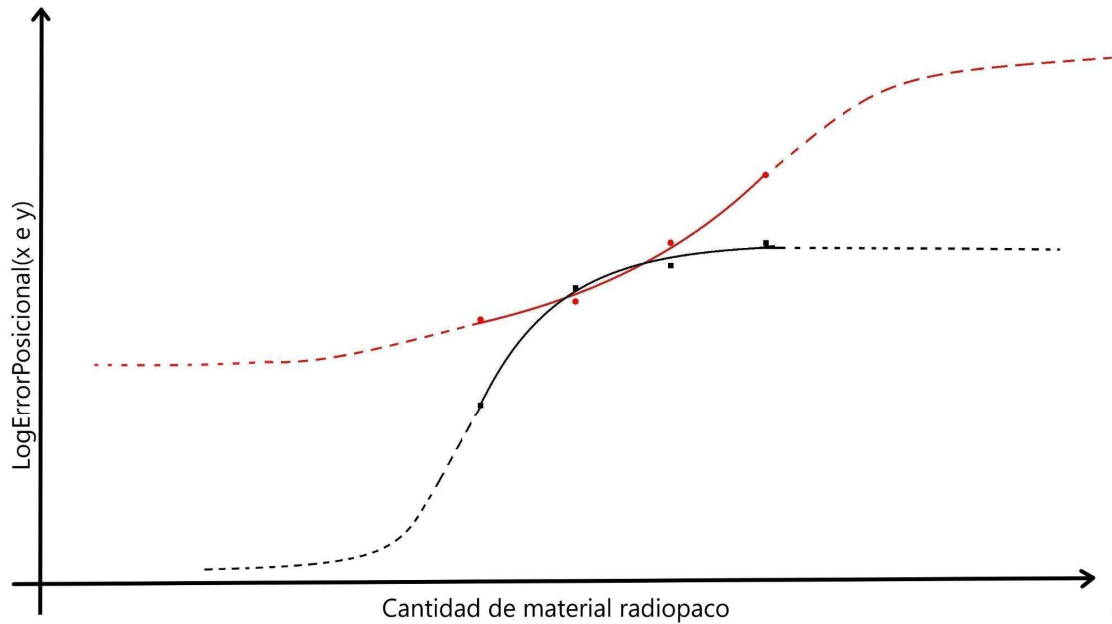


Fig. 17. Gráfica Log10 Error posicional en X vs en Y, con concordancia a teorías de visión y distribución topográfica de fotorreceptores en la retina humana.

Observamos que una de las limitaciones con mayor impacto fue el tamaño de la muestra de los observadores donde fue suficiente para la validación del modelo, aunque un número mayor de observadores de distintas profesiones y con variada experiencia permitirá un análisis más fiable en cuanto a diferencias interobservador. Otro problema fue la metodología de registro de las coordenadas obtenidas, donde algunos de los observadores en primera instancia registraban (anotaban) en papel para luego transcribirlas en la Hoja de Cálculo de Google lo que llevó a un aumento de los errores de registro, en la metodología planteada previamente a la visualización y registro de las coordenadas, se acordó que esta iba a ser de a pares para mitigar este tipo de errores aunque esto no fue posible aumentando los fallos de registro, este problema se puede solucionar en su mayor parte con esta metodología prevista o mediante herramientas de registro automatizado. Finalmente el análisis se limita a mamografía digital 2D indirecta por lo que los resultados obtenidos no podrían extrapolarse a otras modalidades como mamografía directa o tomosíntesis requiriendo nuevos estudios.

Conclusiones

Los resultados obtenidos en este estudio demuestran que es posible avanzar hacia métodos más sensibles y perceptualmente fundamentados para evaluar la calidad de imágenes mamográficas de fantasmas ACR, integrando información cuantitativa basada en la visión humana. Los nuevos métodos de autoría propuestos principalmente el análisis del error posicional permiten caracterizar con mayor precisión el desempeño perceptual de los observadores y la información transmitida por las imágenes, superando las limitaciones de los métodos antecesores que pierden información en el simple hecho del conteo de señales visibles. Los hallazgos sugieren que una fracción significativa de la variabilidad observada puede atribuirse a errores posicionales y, otra fracción a errores de registro, lo que abre la puerta al desarrollo de métricas más ajustadas a las propiedades reales del sistema visual. Este trabajo constituye un aporte exploratorio que sienta las bases conceptuales y metodológicas para futuras investigaciones orientadas a optimizar el control de calidad en mamografía digital con criterios más cercanos al rendimiento perceptual humano.

Conflictos de intereses

Los autores declaran que no tienen conflictos de intereses en relación a este trabajo.

Referencias bibliográficas

1. Berns EA, Pfeiffer DE, Butler PF, Adent C, Baird R, Baker JA, et al. Radiologist's Section Radiologic Technologist's Section Medical Physicist's Section.
2. Huda W, Sajewicz AM, Ogden KM, Scalzetti EM, Dance DR. How good is the ACR accreditation phantom for assessing image quality in digital mammography? Acad Radiol. julio de 2002;9(7):764-72.
3. Historia y estado actual sobre los tipos de procedimientos quirúrgicos realizados en cáncer de mama [Internet]. [citado 29 de abril de 2025]. Disponible en: <http://www.elsevier.es/es-revista-gaceta-mexicana-oncologia-305-pdf-X1665920114361773>
4. Barrón-Gallardo CA, Jave-Suarez LF, Aguilar-Lemarroy A. Historia del cáncer de mama. Rev Médica Inst Mex Seguro Soc. 2020;58(1):75-82.
5. Caicedo Mallarino JJ. Perspectiva histórica de la enfermedad mamaria y su tratamiento. Rev Colomb Obstet Ginecol. 30 de diciembre de 1994;45(4):279-80.
6. Hellman S. Promise, promise, paradigm and prophesy. Nat Clin Pract Oncol. julio de 2005;2(7):325-325.
7. Rando K, Satdjian JL, Alfonso A. Autoridades del Ministerio de Salud Pública.
8. Kundel HL, Nodine CF. Interpreting Chest Radiographs without Visual Search. Radiology. septiembre de 1975;116(3):527-32.
9. Febles G. Screening mamográfico del cáncer de mama, beneficios y controversias. Medicas UIS. abril de 2014;27(1):51-8.
10. Breslau ES. THE CONTINUUM: SOMATIC DISTRESS TO MEDICALIZATION IN WOMEN WITH BREAST CANCER: THEORETICAL AND EMPIRICAL ASSESSMENT. Adv Gend Res. 2003;7:131-80.
11. Garau M, Musetti C, Alonso R, Barrios E. Estadio, perfil biológico y tendencias de incidencia y mortalidad de cáncer de mama en Uruguay: Análisis por franjas etarias. Rev Médica Urug. 1 de marzo de 2024;40(1):e203.
12. Fechner GT, Boring EG, Howes DH, Adler HE. Elements of psychophysics. Volume 1. New York: Holt, Rinehart and Winston, Inc.; 1966.
13. Tanner Jr. WP, Swets JA. A decision-making theory of visual detection. Psychol Rev. 1954;61(6):401-9.
14. Chakraborty DP. New developments in observer performance methodology in medical imaging. Semin Nucl Med. noviembre de 2011;41(6):401-18.
15. Cao JQ, Surgeoner B, Manna M, Boileau JF, Gelmon KA, Brackstone M, et al. Guidance for Canadian Breast Cancer Practice: National Consensus Recommendations for Clinical Staging of Patients Newly Diagnosed with Breast Cancer. Curr Oncol. 2024;31(11):7226-43.
16. Fisher B. Laboratory and clinical research in breast cancer--a personal adventure: the David A. Karnofsky memorial lecture. Cancer Res. noviembre de 1980;40(11):3863-

74.

17. Kumar V, Abbas AK, Aster JC. Robbins & Cotran Pathologic Basis of Disease [Internet]. Elsevier; 2014. (Robbins Pathology). Disponible en: <https://books.google.com.uy/books?id=5NbsAwAAQBAJ>
18. Last JM. A Dictionary of Public Health [Internet]. Oxford University Press; 2007. (Oxford Reference). Disponible en: https://books.google.com.uy/books?id=k3Yd_koxD9gC
19. Bunge MA. Medical Philosophy: Conceptual Issues In Medicine [Internet]. World Scientific Publishing Company; 2013. Disponible en: <https://books.google.com.uy/books?id=DCo8DQAAQBAJ>
20. Schaffner KF. Discovery and Explanation in Biology and Medicine [Internet]. University of Chicago Press; 1993. (Science and Its Conceptual Foundations series). Disponible en: <https://books.google.com.uy/books?id=aVylWcRmne8C>
21. Virchow R. Die Cellularpathologie in ihrer Begründung auf physiologische und pathologische Gewebelehre [Internet]. Philipps-Universität Marburg; 2010. Disponible en: [https://dfg-viewer.de/show/?tx_dlf\[id\]=https%3A%2F%2Farchiv.ub.uni-marburg.de%2Feb%2F2010%2F0274%2Fmets-3087.xml](https://dfg-viewer.de/show/?tx_dlf[id]=https%3A%2F%2Farchiv.ub.uni-marburg.de%2Feb%2F2010%2F0274%2Fmets-3087.xml)
22. Halsted WS. I. The Results of Operations for the Cure of Cancer of the Breast Performed at the Johns Hopkins Hospital from June, 1889, to January, 1894. Ann Surg. noviembre de 1894;20(5):497-555.
23. von Fournier D, Kubli F. Surgical Treatment. En: Kubli F, Bauer M, Kaufmann M, von Fournier D, Junkermann H, editores. Breast Diseases. Berlin, Heidelberg: Springer Berlin Heidelberg; 1989. p. 201-8.
24. PATEY DH, DYSON WH. The prognosis of carcinoma of the breast in relation to the type of operation performed. Br J Cancer. marzo de 1948;2(1):7-13.
25. Madden JL. Modified radical mastectomy. Surg Gynecol Obstet. diciembre de 1965;121(6):1221-30.
26. Fisher Bernard, Anderson Stewart, Bryant John, Margoless Richard G., Deutsch Melvin, Fisher Edwin R., et al. Twenty-Year Follow-up of a Randomized Trial Comparing Total Mastectomy, Lumpectomy, and Lumpectomy plus Irradiation for the Treatment of Invasive Breast Cancer. N Engl J Med. 347(16):1233-41.
27. Veronesi U, Saccozzi R, Del Vecchio M, Banfi A, Clemente C, De Lena M, et al. Comparing radical mastectomy with quadrantectomy, axillary dissection, and radiotherapy in patients with small cancers of the breast. N Engl J Med. 2 de julio de 1981;305(1):6-11.
28. Keynes G. THE PLACE OF RADIUM IN THE TREATMENT OF CANCER OF THE BREAST. Ann Surg. octubre de 1937;106(4):619-30.
29. Mustakallio S. Conservative treatment of breast carcinoma—review of 25 years follow up. Clin Radiol. 1 de enero de 1972;23(1):110-6.
30. Tot T. DCIS, cytokeratins, and the theory of the sick lobe. Virchows Arch Int J Pathol. julio de 2005;447(1):1-8.

31. Tan MPC, Tot T. The sick lobe hypothesis, field cancerisation and the new era of precision breast surgery. *Gland Surg.* 2018;7 6:611-8.
32. Hellman S. Karnofsky Memorial Lecture. Natural history of small breast cancers. *J Clin Oncol Off J Am Soc Clin Oncol.* octubre de 1994;12(10):2229-34.
33. Niibe Y, Hayakawa K. Oligometastases and oligo-recurrence: the new era of cancer therapy. *Jpn J Clin Oncol.* febrero de 2010;40(2):107-11.
34. Palma DA, Olson R, Harrow S, Gaede S, Louie AV, Haasbeek C, et al. Stereotactic ablative radiotherapy versus standard of care palliative treatment in patients with oligometastatic cancers (SABR-COMET): a randomised, phase 2, open-label trial. *The Lancet.* 18 de mayo de 2019;393(10185):2051-8.
35. Ofri A, Moore K. Occult breast cancer: Where are we at? *Breast Edinb Scotl.* diciembre de 2020;54:211-5.
36. Wang R, Yang H xin, Chen J, Huang J jun, Lv Q.
37. Montella L, Riccio V, Ruocco R, Di Marino L, Ambrosino A, Capuozzo A, et al. Occult primary breast cancer and cognates: Atypical today's cases inside a centenarian history. *Curr Probl Cancer Case Rep.* 1 de diciembre de 2022;8:100191.
38. Dialnet-LasTeoriasDeLaPercepcionVisualYElProblemaDelMovimi-3043125 (1).pdf.
39. Curcio CA, Sloan KR, Kalina RE, Hendrickson AE. Human photoreceptor topography. *J Comp Neurol.* 22 de febrero de 1990;292(4):497-523.
40. Carrasco M, P.Talgar C, Cameron EL. Characterizing visual performance fields: effects of transient covert attention, spatial frequency, eccentricity, task and set size. *Spat Vis.* 2001;15(1):61-75.
41. Hacker PMS. Helmholtz's theory of perception: An investigation into its conceptual framework. *Int Stud Philos Sci.* 1995;9(3).
42. Gibson JJ. *The Ecological Approach to Visual Perception.*
43. Zhang A. Human visual detection and search in natural backgrounds. 2024.
44. Wolfe JM. Guided Search 6.0: An updated model of visual search. *Psychon Bull Rev.* 1 de agosto de 2021;28(4):1060-92.
45. Signals_and_Systems_2nd_Edition_by_Oppen.
46. Barrett HH, Swindell W. *Radiological imaging : the theory of image formation, detection, and processing.* New York: Academic Press; 1981.
47. Ley de Weber-Fechner. En: Wikipedia, la enciclopedia libre [Internet]. 2023 [citado 16 de noviembre de 2025]. Disponible en: https://es.wikipedia.org/w/index.php?title=Ley_de_Weber-Fechner&oldid=155120193
48. Green DM, Swets JA. *Signal Detection Theory and Psychophysics* [Internet]. Wiley; 1966. Disponible en: <https://books.google.com.uy/books?id=E-04zQEACAAJ>
49. Swets JA. ROC analysis applied to the evaluation of medical imaging techniques. *Invest Radiol.* abril de 1979;14(2):109-21.

50. Chakraborty DP, Berbaum KS. Observer studies involving detection and localization: modeling, analysis, and validation. *Med Phys.* agosto de 2004;31(8):2313-30.
51. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. *J Thorac Oncol Off Publ Int Assoc Study Lung Cancer.* septiembre de 2010;5(9):1315-6.
52. Thompson JD, Manning DJ, Hogg P. The Value of Observer Performance Studies in Dose Optimization: A Focus on Free-Response Receiver Operating Characteristic Methods. *J Nucl Med Technol.* 1 de junio de 2013;41(2):57.
53. Philip C. Bunch, John F. Hamilton, Gary K. Sanderson, Arthur H. Simmons. A Free Response Approach To The Measurement And Characterization Of Radiographic Observer Performance. En 1977. p. 124-35. Disponible en: <https://doi.org/10.1117/12.955926>
54. [\[https://biblus.us.es/bibing/proyectos/use/abreproy/10549/fichero/cap%C3%ADtulo2.pdf\]](https://biblus.us.es/bibing/proyectos/use/abreproy/10549/fichero/cap%C3%ADtulo2.pdf)(<https://biblus.us.es/bibing/proyectos/use/abreproy/10549/fichero/cap%C3%ADtulo2.pdf>).
55. Mandelson MT, Oestreicher N, Porter PL, White D, Finder CA, Taplin SH, et al. Breast density as a predictor of mammographic detection: comparison of interval- and screen-detected cancers. *J Natl Cancer Inst.* 5 de julio de 2000;92(13):1081-7.
56. Duffy SW, Tabár L, Yen AMF, Dean PB, Smith RA, Jonsson H, et al. Mammography screening reduces rates of advanced and fatal breast cancers: Results in 549,091 women. *Cancer.* 1 de julio de 2020;126(13):2971-9.
57. Gosling SB, Arnold EL, Davies SK, Cross H, Bouybayoune I, Calabrese D, et al. Microcalcification crystallography as a potential marker of DCIS recurrence. *Sci Rep.* 8 de junio de 2023;13(1):9331.
58. Krupinski EA. Visual scanning patterns of radiologists searching mammograms. *Acad Radiol.* febrero de 1996;3(2):137-44.
59. Shen Y, Dai W, Richards VM. A MATLAB toolbox for the efficient estimation of the psychometric function using the updated maximum-likelihood adaptive procedure. *Behav Res Methods.* marzo de 2015;47(1):13-26.
60. Dromain C, Boyer B, Ferré R, Canale S, Delalogue S, Balleyguier C. Computed-aided diagnosis (CAD) in the detection of breast cancer. *Eur J Radiol.* marzo de 2013;82(3):417-23.
61. C. E. Shannon. A mathematical theory of communication. *Bell Syst Tech J.* julio de 1948;27(3):379-423.
62. Luce1964.
63. Sundell VM, Mäkelä T, Meaney A, Kaasalainen T, Savolainen S. Automated daily quality control analysis for mammography in a multi-unit imaging center. *Acta Radiol.* febrero de 2019;60(2):140-8.
64. [\[https://www.cirsinc.com/wp-content/uploads/2019/04/015-DS-120418.pdf\]](https://www.cirsinc.com/wp-content/uploads/2019/04/015-DS-120418.pdf)(<https://www.cirsinc.com/wp-content/uploads/2019/04/015-DS-120418.pdf>).

65. Cochran G, Cox M. Experimental designs.
66. María Sarasua, Horacio Botti Alsina, Federico Lecumberry. Evaluación por expertos y aprendizaje automático de la calidad de imágenes mamográficas y del proceso enseñanza aprendizaje en la Licenciatura en Imagenología de la EUTM. EUTM, Facultad de Medicina; 2019.
67. Manual de usuario Barco Nio 5MP LED (MDNG-5221) (55 páginas) [Internet]. [citado 23 de octubre de 2025]. Disponible en: <https://www.manual.ec/barco/nio-5mp-led-mdng-5221/manual?p=8>
68. Neyman J, Pearson ES. On the problems of the most efficient tests of statistical hypotheses. Philos Trans R Soc Lond. 1933;231A:289-338.
69. Macmillan NA, Creelman CDouglas. Detection theory : a user's guide. 2nd ed. Mahwah, N.J: Lawrence Erlbaum Associates; 2005.
70. Barufaldi B, Lau KC, Schiabel H, Maidment DA. Computational assessment of mammography accreditation phantom images and correlation with human observer analysis. En 2015. p. 941606. Disponible en: <https://ui.adsabs.harvard.edu/abs/2015SPIE.9416E..06B>
71. Brooks KW, Trueblood JH, Kearfott KJ. Subjective evaluations of mammographic accreditation phantom images by three observer groups. Invest Radiol. enero de 1994;29(1):42-7.
72. Dougherty G. Computerized evaluation of mammographic image quality using phantom images. Comput Med Imaging Graph. 1 de septiembre de 1998;22(5):365-73.

Anexos

1. [Registro de datos original](#)
2. [Histogramas estimación del valor crítico \(SD\) e interpretación de errores de registro.](#)
3. SET DE DATOS OBTENIDOS (TABLAS): [Set de datos con errores de registro identificados](#)
4. SCRIPT PARA R: [Script de desvios criticos](#)

