

ACTUALIZACIÓN

¿Son los sistemas de inteligencia artificial una herramienta útil para los programas de cribado de cáncer de mama?

O. Díaz^a, A. Rodríguez-Ruiz^b, A. Gubern-Mérida^b, R. Martí^c y M. Chevalier^{d,e,*}^a Departamento de Matemáticas e Informática, Universidad de Barcelona, Barcelona, España^b ScreenPoint Medical BV, Nijmegen, Países Bajos^c Instituto de Visión Artificial y Robótica (VICOROB), Universitat de Girona, Girona, España^d Física Médica, Departamento de Radiología, Rehabilitación y Fisioterapia, Universidad Complutense de Madrid, Madrid, España^e Instituto de Investigación Sanitaria, Hospital Clínico San Carlos, Madrid, España

Recibido el 16 de agosto de 2020; aceptado el 16 de noviembre de 2020

Disponible en Internet el 15 de enero de 2021

PALABRAS CLAVEMamografía;
Inteligencia artificial;
Aprendizaje profundo;
CAD;
Cribado de cáncer de mama

Resumen Los programas poblacionales de detección precoz del cáncer de mama (PDPCM) contribuyen eficazmente a la reducción de la mortalidad debida a esta patología. Estos programas usan la mamografía para el seguimiento de las mujeres invitadas a participar. La mamografía digital posibilitó el desarrollo de sistemas de diagnóstico asistido por ordenador (CAD) con importantes expectativas para reducir la carga de trabajo de los radiólogos participantes en dichos programas. Distintas investigaciones evidenciaron la obtención de un número importante de falsos positivos con los CAD. La creciente implementación de sistemas basados en inteligencia artificial ha propiciado el desarrollo de investigaciones sobre su rendimiento en mamografía que muestran su superioridad frente a los CAD tradicionales. En este artículo se describen los fundamentos de los sistemas de inteligencia artificial y se aporta una panorámica de sus posibles aplicaciones en el marco de los programas de cribado de mama.

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

KEYWORDSMammography;
Artificial intelligence;
Deep learning;
CAD;
Breast cancer screening**Are artificial intelligence systems useful in breast cancer screening programs?**

Abstract Population-based breast cancer screening programs are efficacious in reducing the mortality due to breast cancer. These programs use mammography to screen the women who are invited to participate. Digital mammography makes it possible to develop computer-assisted diagnosis (CAD) systems that promise to reduce the workload of radiologists participating in screening programs. However, various studies have shown that CAD results in a high rate of false

* Autor para correspondencia.

Correo electrónico: chevalie@ucm.es (M. Chevalier).

positive diagnoses. Systems based on artificial intelligence are being more widely implemented, and studies have shown that these systems have better diagnostic performance than traditional CAD systems.

This article explains the fundamentals of artificial intelligence systems and an overview of possible applications of these systems within the framework of breast cancer screening programs.

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

Introducción

El cáncer de mama es el tumor maligno más frecuente entre las mujeres con una incidencia anual de más de 2 millones de casos en todo el mundo y una tasa de mortalidad del 30%¹. En España se han estimado 32 800 nuevos casos en el año 2020, con una tasa de mortalidad del 19,5%. La detección precoz del cáncer de mama es vital para disminuir la mortalidad^{2,3} y mejorar el pronóstico y la calidad de vida de la mujer tras el tratamiento. Desde hace décadas, en la mayoría de los países se organizan programas de cribado o detección precoz del cáncer de mama (PDPCM)^{4,5} dirigidos a mujeres asintomáticas. La mamografía es la prueba diagnóstica en la que se basan estos programas debido a su alta sensibilidad para la detección del cáncer, así como a un balance coste-beneficio adecuado. La población diana en la mayoría de los PDPCM son mujeres con edades comprendidas entre 50 y 69 años a las que se les suele realizar un examen mamográfico cada 2 años que, en la mayoría de los casos, comprende dos proyecciones por mama. La efectividad de los PDPCM está limitada por distintos factores, siendo el principal el empleo de una técnica 2D como es la mamografía, que genera solapamiento de tejidos y reduce la capacidad de detección⁶⁻⁸. Otros factores relevantes son la alta carga de trabajo, derivada del elevado número de mamografías (la mayoría normales) y del procedimiento de doble lectura⁹. Este procedimiento, adoptado para aumentar la tasa de detección de cáncer, se ve últimamente obstaculizado por la falta creciente de radiólogos^{10,11}. Estas limitaciones influyen, entre otros aspectos, en la tasa de cánceres de intervalo, que son cánceres diagnosticados de forma sintomática tras un examen de cribado negativo y antes de la siguiente ronda. Según los estudios, estos cánceres son alrededor de 2 por 1000 cribados en los PDPCM europeos⁷. El empleo de la tomosíntesis de mama ha supuesto un incremento notable en la tasa de detección de cáncer de mama¹²⁻¹⁴. Sin embargo, su inclusión en los PDPCM está cuestionada al incrementar la carga de trabajo y el tiempo de lectura^{15,16}. Los sistemas de detección asistida por ordenador (CAD, por sus siglas en inglés de *computer-aided detection*) aparecen en los años 90 como una herramienta de apoyo al diagnóstico en mamografía. Estos sistemas procesan automáticamente las imágenes y señalan las áreas consideradas sospechosas (fig. 1) para reducir el número de lesiones que no son visualizadas. Existen dos categorías de algoritmos de CAD: los utilizados para detectar la

presencia de una lesión y los que, además, indican si la lesión es benigna o maligna. En ambos casos, los algoritmos “buscan” en la imagen características específicas de las lesiones previamente definidas por los radiólogos. El inconveniente mayor de los CAD es su moderada especificidad, marcando muchas áreas como sospechosas que conducen a falsos positivos o negativos de los radiólogos¹⁷⁻²⁰. Los desarrollos recientes habidos en el campo de la inteligencia artificial (IA) abren un nuevo horizonte en el terreno de los sistemas de apoyo al diagnóstico. Los resultados de numerosas investigaciones evidencian que los sistemas basados en el “aprendizaje profundo” [*deep learning* (DL)] tienen un rendimiento mayor que el CAD tradicional.

En este artículo se trata de proporcionar una panorámica de las aplicaciones de los sistemas de IA para la detección del cáncer de mama y sus funciones en el entorno de los PDPCM. Previamente, se abordan los conceptos básicos más relevantes en los que se fundamentan los algoritmos de IA.

Inteligencia artificial: evolución de las técnicas de aprendizaje automático

Conceptos tales como DL o “aprendizaje automático” [*machine learning* (ML)] son habituales hoy en día en ámbitos muy diversos incluida la radiología²¹. Su aplicación en el diagnóstico por imagen supone una auténtica revolución como evidencia el importante aumento del número de publicaciones y sesiones científicas dedicadas a este tema. La IA estudia la incorporación de comportamientos de la inteligencia humana a las máquinas como, por ejemplo, el aprendizaje. Dentro de la IA encontramos entre otros la robótica o el ML (fig. 2) que es el elemento que provee a las máquinas de la capacidad de aprender una tarea a partir de experiencias previas sin necesidad de una programación específica. Por otro lado, el DL^{22,23} comprende técnicas de aprendizaje automático en las que el algoritmo aprende por sí solo las características más importantes de una imagen para realizar una tarea determinada. Las arquitecturas de aprendizaje más utilizadas actualmente en DL son las *redes neuronales*, que corresponden a nodos interconectados formando múltiples capas que imitan redes neuronales biológicas cerebrales. Una vez entrenadas, las redes asignan un peso a cada nodo (característica de la imagen) para destacar la importancia de este en la tarea final de la red (segmentación, detección, etc.).

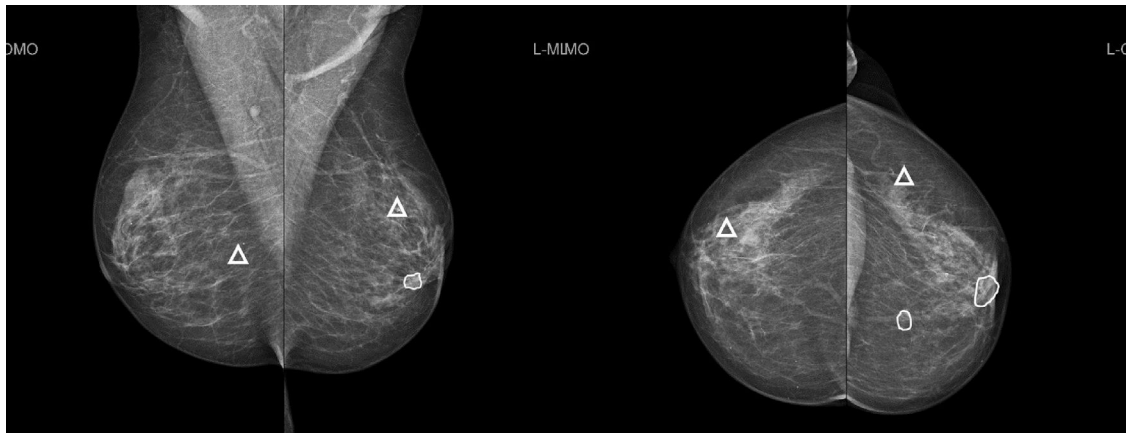


Figura 1 Ejemplo de un sistema tradicional de diagnóstico asistido por ordenador. Automáticamente procesa las imágenes y muestra al lector marcas en diversas áreas consideradas sospechosas por el sistema.

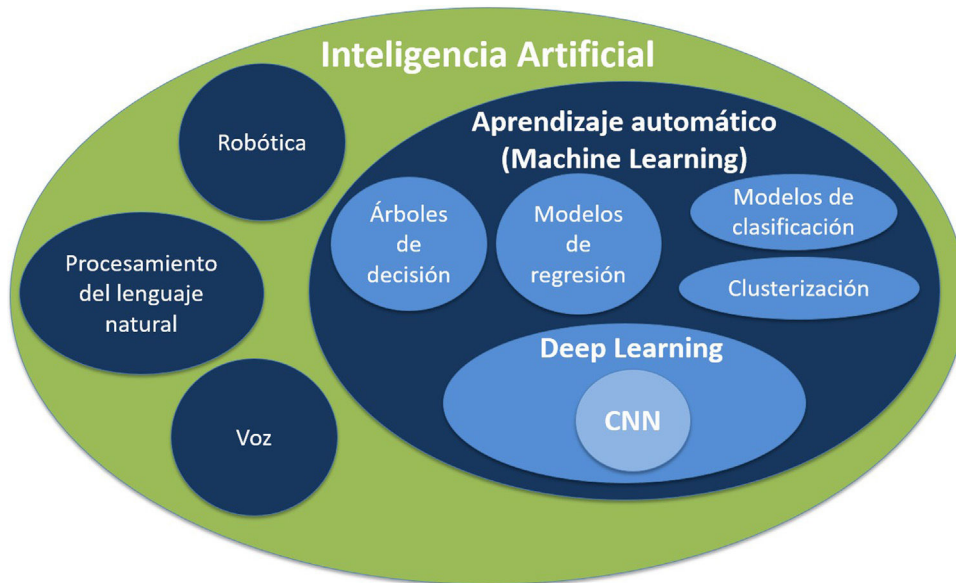


Figura 2 Diagrama explicativo de las áreas que abarca la inteligencia artificial. La sigla CNN es el acrónimo del término inglés “convolutional neural networks” (en español: redes neuronales convolucionales).

Dentro del DL, las redes neuronales convolucionales o CNN (del inglés *convolutional neural networks*) son un subconjunto de algoritmos basados en operaciones de *convolución* que comenzaron a utilizarse a principios de los años 90²⁴. En 1996 se publicó el primer trabajo científico en el que se utilizó una CNN aplicada a mamografía²⁵. No ha sido hasta esta década cuando esta técnica se ha explotado masivamente impulsada por la disponibilidad de gran cantidad de imágenes médicas digitales, la mejora de la potencia de cálculo y el abaratamiento de las unidades de procesamiento gráfico.

Las CNN aprenden de forma automática la combinación de las propiedades de la imagen que posibilitan hacer una tarea determinada sin necesidad de un conocimiento previo ni, por tanto, de la intervención humana. Esto constituye la diferencia fundamental entre las CNN y los CAD tradicionales. Las CNN entrenadas con un gran número de casos tienen un rendimiento muy superior a los CAD

y evitan los sesgos asociados al uso de características escogidas manualmente²⁶.

Conceptos básicos en inteligencia artificial

Dentro del ML se distinguen dos grandes familias de algoritmos: el *aprendizaje supervisado* y el *aprendizaje no supervisado*. En los primeros, una tarea se aprende a partir de un conjunto de casos con resultado conocido (*ground truth*). En el ámbito radiológico son imágenes con una etiqueta conocida dependiendo del problema que se pretende abordar: localización de una lesión (*detección*), delimitación del borde de la lesión (*segmentación*) o diagnóstico de la lesión (*clasificación*). Los modelos de aprendizaje no supervisado utilizan imágenes sin etiquetar que son agrupadas de acuerdo con unas características comunes. La carencia de un número elevado de imágenes etiquetadas representa uno de los grandes problemas actuales para los modelos de

aprendizaje supervisado. Esto ha motivado una investigación creciente de modelos no supervisados y la propuesta de algoritmos que combinan ambos paradigmas. Actualmente, existen *bases de imágenes públicas* y etiquetadas, siendo las más comunes en mamografía digital las INbreast²⁷ y CBIS-DDSM²⁸. También existen bases privadas como OMI-DB²⁹. Muchas arquitecturas de DL requieren una imagen como elemento de entrada, mientras que otras trabajan a partir de trozos o regiones de interés, lo que permite aumentar el peso de las imágenes sin necesidad de disponer de una gran capacidad de computación.

El aprendizaje en los algoritmos supervisados se divide en tres fases:

1. Entrenamiento (*training*): los parámetros del modelo se optimizan en función de un coste (error) previamente determinado que se trata de minimizar. El error se obtiene al comparar la predicción del modelo con la etiqueta real asociada al conjunto de imágenes de entrenamiento.
2. Validación (*validation*): el modelo obtenido tras el entrenamiento se usa para predecir las etiquetas de un segundo conjunto de imágenes. Los resultados obtenidos con este conjunto sirven también para hacer una primera validación del modelo y ajustar parámetros y características para optimizarlo.
3. Evaluación (*test*): se evalúa el rendimiento del modelo final con un conjunto de nuevas imágenes no vistas anteriormente. Los resultados derivados del análisis con este conjunto determinan la precisión y capacidad de generalización del modelo.

Los conjuntos de imágenes utilizados para la validación y evaluación deben tener una distribución de casos lo más parecida posible a la distribución de los que evaluará el algoritmo en su aplicación final (imágenes de equipos de diferentes fabricantes, proporción de casos con lesiones observadas en cribado poblacional, etc.).

El tamaño reducido de las bases de datos en radiología es un problema importante que se trata de paliar incrementando el conjunto de imágenes de entrenamiento (*data augmentation*) aplicando diferentes técnicas. La más popular consiste en deformar las imágenes, siendo la más empleada la rotación. Otras técnicas, como la conocida por *transfer learning*, utilizan imágenes cotidianas para el entrenamiento del modelo de DL aprovechando la existencia de características comunes en imágenes de diferentes dominios. En su aplicación a mamografía ha supuesto una ligera mejora en el rendimiento del DL³⁰. Más recientemente, se han introducido redes neuronales capaces de generar imágenes sintéticas y enriquecer así el conjunto de entrenamiento (GAN³¹, del inglés *Generative Adversarial Networks*)³². Estas redes han sido propuestas originalmente como algoritmos de aprendizaje no supervisado y su uso se ha extendido a los modelos de aprendizaje semisupervisado.

Las CNN procesan las imágenes a través de múltiples etapas secuenciales denominadas capas en las que se aplican diferentes procedimientos matemáticos. Las imágenes son primeramente preprocesadas, normalizando su intensidad para independizar el aprendizaje de la red del origen de las imágenes (distintos equipamientos, etc.). Las etapas siguientes involucran diferentes procedimientos; los más

comunes son las convoluciones, la reducción de dimensiones (*maxpooling*) y la renormalización (*softmax*) entre otras. Existen CNN con arquitecturas distintas que difieren en aspectos tales como el número de capas, las funciones de activación, etc.

Algoritmos de inteligencia artificial aplicados a la imagen médica

La falta de grandes bases de datos públicas ha limitado la aplicación clínica de la IA en imagen médica a pesar de la gran fuente de información que constituyen los sistemas actuales de archivado de imágenes o PACS (del inglés *Picture Archiving and Communication System*). Aun así, se han desarrollado una gran cantidad de aplicaciones muy útiles para el diagnóstico en general y, en concreto, para el diagnóstico de patologías mamarias.

La tarea más popular de las aplicaciones de IA es la detección automática de lesiones en las diferentes modalidades de imagen. Esta tarea conlleva la localización de regiones en la imagen con una alta probabilidad de ser una lesión, pudiendo llegar a detectarse diferentes tipos de lesiones según el entrenamiento que haya recibido el algoritmo. Recientemente, se ha demostrado que las CNN pueden alcanzar en mamografía las tasas de detección de un radiólogo medio³³. Una vez detectada la lesión, la CNN puede clasificarla en maligna o benigna o identificar a qué tipo de lesión pertenece (nódulo, microcalcificación, etc.). La clasificación hecha por la CNN tiene la ventaja de eliminar la variabilidad que puede existir entre observadores. Otra de las aplicaciones más comunes de las CNN es la segmentación de tejidos o lesiones (delimitación de su contorno)^{34,35}. Para este objetivo, las redes más populares son las U-Net que permiten diferenciar tejidos (tejido glandular, adiposo o músculo pectoral en imágenes mamográficas) o calcular el volumen de una lesión³⁵. El registro de imágenes se utiliza para analizar los cambios en imágenes de un mismo paciente obtenidas en distintos momentos (registro longitudinal) o usando distintas modalidades de imagen (registro multimodal). Esta es una tarea compleja para las CNN al tratar de simular la forma en que los humanos perciben los cambios en una imagen, si bien los resultados con ellas obtenidos muestran un incremento en la velocidad y precisión de esta tarea³⁶.

Inteligencia artificial como herramienta para mejorar el cribado de cáncer de mama

Los recientes avances en IA abren posibilidades que van más allá de las aportadas por el CAD tradicional en el apoyo a los radiólogos durante la lectura de los exámenes de mamografía en los PDPCM. Los objetivos de los sistemas de IA son los mismos que los del CAD: mejorar la detección de lesiones malignas, reducir cánceres de intervalo y, al mismo tiempo, reducir la carga de lectura. Eventualmente, es posible que los nuevos sistemas y técnicas de IA puedan incluso mejorar la ratio entre coste y beneficio de los PDPCM³⁷.

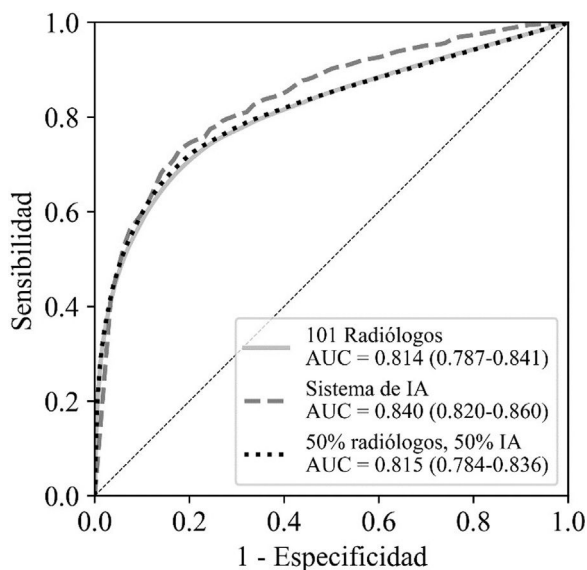


Figura 3 Curvas ROC (*Receiver Operating Characteristics*) obtenidas en un estudio con 101 radiólogos y 2600 casos, 650 de ellos con cáncer. Se representan la curva ROC promedio de los 101 radiólogos (línea continua), la obtenida con un sistema de inteligencia artificial (curva a trazos) y la curva ROC obtenida en un escenario donde el 50% de los casos son leídos por el sistema de inteligencia artificial y el otro 50% por los 101 radiólogos (curva a puntos). El parámetro AUC representa el área bajo la curva ROC (del inglés *Area Under the Curve*). Cortesía de A. Rodríguez-Ruiz³³.

Rendimiento de los sistemas de inteligencia artificial en la detección de cáncer en mamografía

La primera pregunta que cabe hacerse es si los sistemas de IA son suficientemente precisos en la detección de cáncer en mamografía. Es de esperar que así sea dado que estos sistemas basan su funcionamiento en técnicas de aprendizaje automático y son entrenados con bases de datos que incluyen miles de imágenes de mamografía con una casuística muy variada. Las investigaciones realizadas indican que la precisión de estos sistemas supera a la de los CAD tradicionales³⁸. Kim et al.³⁹ estudiaron el rendimiento de un sistema de IA utilizando exámenes de una población de cribado (1200 exámenes, la mitad con cáncer), obteniendo una sensibilidad del 76% y una especificidad del 89% que es mucho mayor que la de los CAD tradicionales (alrededor del 50%). Los estudios en los que se compara directamente el rendimiento del sistema de IA y el de los radiólogos muestran un rendimiento inferior o similar al de los radiólogos utilizando muestras enriquecidas⁴⁰ o exámenes de cribado⁴¹. Más recientemente, se han publicado estudios en los que se compara un mismo sistema de IA frente a 14 radiólogos utilizando una muestra de 240 casos (100 con cáncer)⁴² y 101 radiólogos y una muestra de 2600 casos (650 aproximadamente con cáncer)³³. En ambos casos, el sistema de IA obtuvo un rendimiento estadísticamente similar al promedio de todos los radiólogos en términos del área bajo la curva (AUC) de la curva ROC (característica operativa del receptor) (fig. 3).

El éxito de un PDPCM depende de forma importante de la uniformidad en el rendimiento de los radiólogos involu-

crados. Los sistemas de IA vendrían a paliar las diferencias notables en sensibilidad existentes entre profesionales que, según un estudio reciente, están entre el 53% y el 89% y varían en función del tipo de lesión⁴³.

Aunque los resultados son prometedores, todavía sería posible mejorar el rendimiento de los algoritmos de IA para alcanzar como mínimo el nivel del mejor lector humano con una precisión independiente de la calidad de imagen de las mamografías producidas por equipos de distintos fabricantes^{44,45}. Esto exige una innovación continua de los algoritmos para poder emular los procedimientos complejos que siguen los radiólogos durante la lectura rutinaria de los exámenes de mamografía tales como las comparaciones con la mama contralateral o con los exámenes de la ronda de cribado anterior.

Inteligencia artificial como instrumento de apoyo para los radiólogos en los PDPCM

Como ya se ha indicado, los sistemas tradicionales de CAD presentan un rendimiento pobre para la detección de cáncer debido a su baja especificidad; a ello hay que unir la nula interacción entre el radiólogo y el sistema que muestra exclusivamente marcas. Ambos aspectos se traducen en un apoyo insuficiente¹⁷, que puede además generar problemas importantes cuando no se utiliza de forma conveniente¹⁸. La nueva generación de CAD basados en IA para mamografía persigue evitar que hasta un tercio de las lesiones malignas no sean detectadas⁴⁶ durante la lectura de las mamografías del cribado. Para ello, además de mostrar marcas en regiones sospechosas (con menos falsos positivos que un CAD tradicional), algunos sistemas introducen nuevos conceptos como el soporte interactivo (fig. 4) o una puntuación a todo el examen indicativa de la sospecha de un cáncer^{42,47}.

El primer estudio publicado sobre el impacto en el rendimiento de los radiólogos debido al uso de un sistema de IA con soporte interactivo, marcas de CAD y una puntuación para todo el examen, demuestra una mejora tanto en sensibilidad como en especificidad de los 14 radiólogos participantes con una muestra de 240 casos, de los cuales 100 eran con cáncer⁴². Aunque el incremento del rendimiento fue moderado (3%), hay que destacar que fue independiente de la experiencia de los radiólogos. Este estudio también evidenció que no aumentó el tiempo de lectura por caso con el sistema de IA a diferencia de lo que se observó en estudios previos empleando un CAD tradicional en el que se encontraron incrementos de hasta un 20% por caso⁴⁸.

Todos estos resultados tan prometedores han de contrastarse con estudios prospectivos hechos con poblaciones de PDPCM para profundizar en el impacto de los sistemas de IA en la mejora del rendimiento de los radiólogos⁴⁹. Este es un paso imprescindible para medir la eficiencia de los algoritmos de IA en un entorno realista.

Inteligencia artificial como lector autónomo en los PDPCM

En los apartados anteriores se han comentado los resultados de las aplicaciones de los sistemas de IA cuando son empleados en paralelo con los radiólogos, es decir, cuando

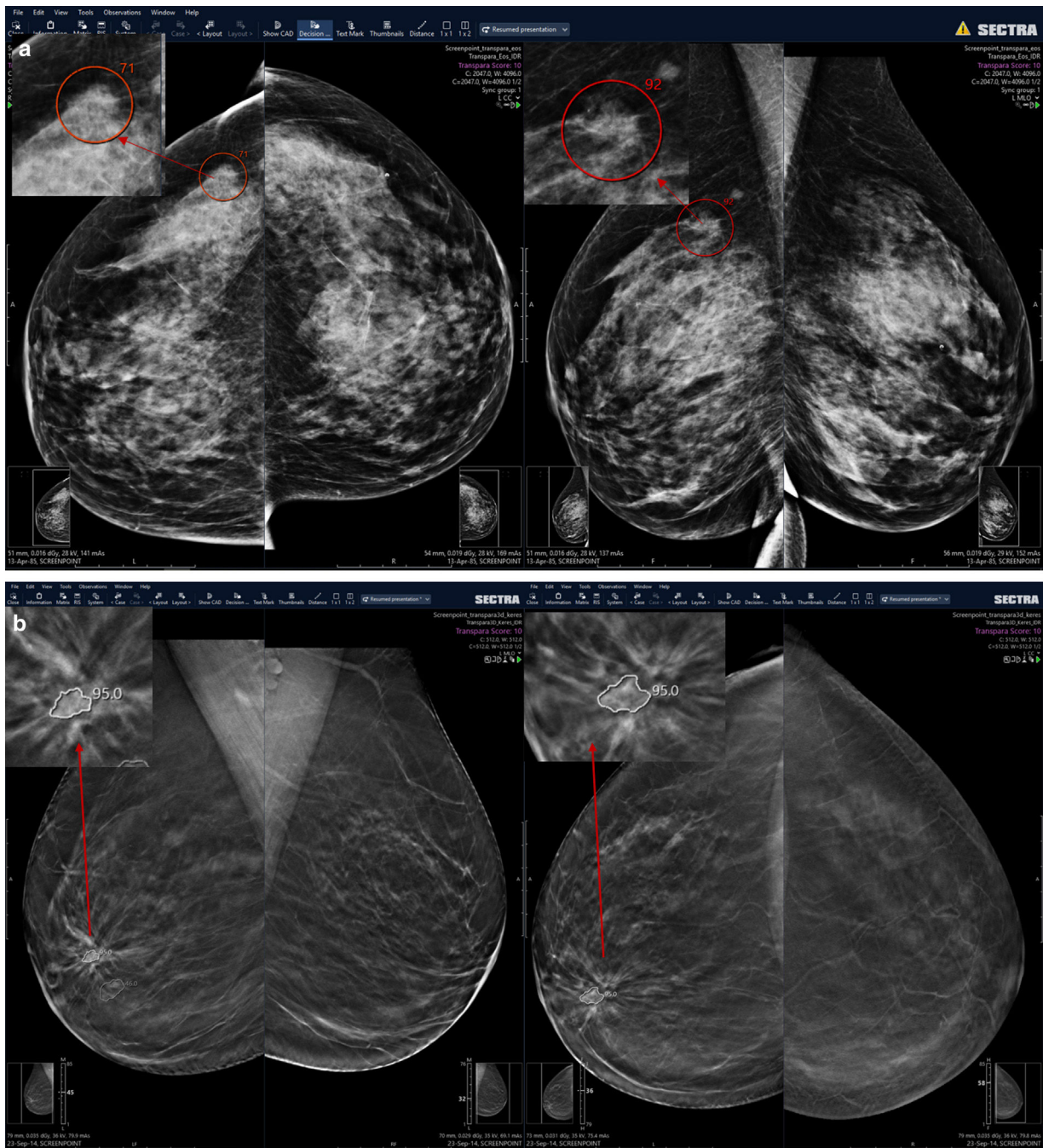


Figura 4 Ejemplo de un sistema de inteligencia artificial proporcionando un soporte interactivo en mamografía más allá de la funcionalidad del diagnóstico asistido por ordenador tradicional. Tras señalar el área de interés, el sistema muestra automáticamente el nivel de sospecha en las dos vistas craneocaudal y mediolateral oblicua tal y como puede apreciarse en los insertos incluidos en los casos (a) grupo de microcalcificaciones y (b) distorsión. Cortesía de N. Karssemeijer. Radboud University Medical Center, Nijmegen, Países Bajos.

su funcionalidad es la de un sistema de CAD. Sin embargo, estos sistemas pueden ir más allá y tener una utilidad como lector autónomo al tener la capacidad de categorizar las mamografías según su probabilidad de contener un cáncer. Aquellos casos clasificados por el algoritmo como más sospechosos serían revisados por los radiólogos en primer lugar, evitando que las lesiones no sean detectadas y favoreciendo que la mujer sea rellamada lo antes posible. Considerando

esta capacidad y que más del 95% de las mujeres en la población de cribado no tienen ningún tipo de anomalía, es posible que uno de los usos futuros más destacados de los sistemas de IA sea como primer lector. Su papel sería clasificar automáticamente los exámenes reduciendo la carga de trabajo al evitar que los exámenes normales tengan que ser leídos por dos radiólogos, en el caso de realizar doble lectura.

El uso de los sistemas de IA como lectores autónomos actuando como primeros, segundos, o incluso terceros lectores de los exámenes de mamografía reduciría significativamente la carga de trabajo de los PDPCM. En este contexto, el reto mayor para un sistema de IA está en proporcionar automáticamente una decisión de rellamada basándose, por ejemplo, en la puntuación asignada al examen una vez probada la alta precisión del sistema^{43,48}. En un reciente estudio retrospectivo⁵⁰, en el que se utilizaron 8000 exámenes de un PDPCM multicéntrico del Reino Unido, el sistema de IA (entrenado con parte de estos datos) se utilizó como lector autónomo de los exámenes con menos probabilidad de cáncer, y el resto fueron leídos por un radiólogo. Los resultados permitieron concluir que el uso de IA podría conducir a una reducción de la carga de trabajo del 42,8%, sin originar un incremento de los falsos positivos o de los falsos negativos del programa. En otro estudio retrospectivo, Rodríguez-Ruiz et al.⁵¹ analizaron la posibilidad de usar un sistema de IA entrenado con datos independientes para leer los casos con menos probabilidad de cáncer de una muestra de cribado con 2600 casos. La muestra contenía 650 casos con cáncer y estaba enriquecida con imágenes de distintos fabricantes. Los investigadores encontraron que con el sistema de IA se reducía en un 47% el número de casos del cribado que debían ser leídos, con una disminución de falsos positivos de un 27%, aunque los falsos negativos se incrementaron en un 7%. Estos resultados expresados de una forma más general indican que reducir en un 20% el número de casos que han de ser leídos en un programa de cribado supone una disminución de los falsos positivos en un 5% y un incremento inferior al 1% de los falsos negativos.

Inteligencia artificial en tomosíntesis y futuras aplicaciones

Ya se ha mencionado el gran potencial de la IA para reducir el tiempo de lectura de los exámenes de tomosíntesis^{15,16}. La tarea del sistema de IA en esta modalidad sería buscar y mostrar al radiólogo de forma automática el corte tomográfico en el que se encuentra la región con la anormalidad. En estudios recientemente publicados se demuestra que se pueden lograr reducciones en el tiempo de lectura por examen entre el 15% y el 50%⁵²⁻⁵⁴. Esta reducción favorecería la inclusión de la tomosíntesis en los PDPCM. En estos estudios también se evidencia la posibilidad de utilizar el sistema de IA para generar una imagen sintética en la cual se realcen las lesiones detectadas en los cortes incrementando de forma notable el rendimiento de estas imágenes⁵⁴.

Más allá de ayudar al radiólogo a detectar cáncer o diferenciar lesiones malignas de benignas, los sistemas de IA podrían también diferenciar lesiones invasivas de otras como carcinomas ductales *in situ* (DCIS de sus siglas en inglés), o diferenciar qué DCIS acabará siendo invasivo^{55,56}. Pese a la dificultad de esta tarea, los sistemas de IA tienen la capacidad de analizar características que pueden estar ocultas a la vista humana en línea con la radiómica. La cuantificación automática del riesgo que pueden representar parámetros tales como la densidad de la mama^{57,58} u otros patrones del tejido glandular⁵⁹ representa otro posible uso avanzado de los sistemas de IA.

Tal y como demuestra el estudio de Wang et al.⁶⁰, el empleo de la IA posibilita también detectar otras patologías en la mamografía como son lesiones vasculares a partir del análisis de las calcificaciones vasculares de la mama con una precisión similar a la de expertos radiólogos.

Conclusiones

Los sistemas de IA proporcionan un rendimiento superior al de los CAD tradicionales con una sensibilidad similar y una reducción muy significativa de los falsos positivos. Estos sistemas posibilitan además categorizar los exámenes mamográficos en función de la probabilidad de contener un cáncer. Esta característica constituye una herramienta importante de cara a reducir el flujo de trabajo de los PDPCM y crea expectativas importantes sobre su uso como primer lector.

Los sistemas de IA desempeñan también un papel relevante de cara a reducir la variabilidad en la sensibilidad de los radiólogos durante las lecturas del alto volumen de mamografías de los PDPCM. Este aspecto abre las puertas a considerar al sistema de IA como un referente de cara a evaluar la efectividad del PDPCM una vez demostrada su capacidad.

Hasta ahora, las investigaciones sobre el rendimiento autónomo de los sistemas de IA se han basado en estudios retrospectivos y, en muchos casos, con muestras enriquecidas. El análisis del funcionamiento independiente de estos sistemas en escenarios más realistas y, en concreto, en entornos asociados a los PDPCM es necesario para evaluar su efectividad real.

Las aplicaciones de los algoritmos de IA en el marco de la imagen médica han de estar sustentados en aspectos éticos y médico-legales que han de ser discutidos y quedar bien establecidos. Uno de los problemas más destacados que han de abordarse es el de la responsabilidad ante un posible fallo del sistema.

Autoría

1. Responsable de la integridad del estudio: OD, ARR, AGM, RM, MCH.
2. Concepción del estudio: OD, ARR, AGM, RM, MCH.
3. Diseño del estudio: OD, ARR, AGM, RM, MCH.
4. Obtención de los datos: OD, ARR, AGM, RM, MCH.
5. Análisis e interpretación de los datos: OD, ARR, AGM, RM, MCH.
6. Tratamiento estadístico: OD, ARR, AGM, RM, MCH.
7. Búsqueda bibliográfica: OD, ARR, AGM, RM, MCH.
8. Redacción del trabajo: OD, ARR, AGM, RM, MCH.
9. Revisión crítica del manuscrito con aportaciones intelectualmente relevantes: OD, ARR, AGM, RM, MCH.
10. Aprobación de la versión final: OD, ARR, AGM, RM, MCH.

Conflicto de intereses

R. Martí, O. Díaz y M. Chevalier forman parte del equipo investigador del proyecto ICEBERG ("Computación de la imagen para la mejora de la radiómica del cáncer de Mama" [Ref. RTI2018-096333-B-I00]) financiado por el Ministerio de

Ciencia, Innovación y Universidades. A. Rodríguez Ruiz y A. Gubern Merida son empleados de ScreenPoint Medical BV, Nijmegen, Países Bajos.

Bibliografía

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68:394–424.
- Sociedad Española de Oncología Médica (SEOM). Disponible en: https://seom.org/seomcms/images/stories/recursos/Cifras_del_cancer.2020.pdf; 2020.[consultada 31/07/2020].
- Sardanelli F, Aase HS, Álvarez M, Azavedo E, Baarslag HJ, Balleyguier C, et al. Position paper on screening for breast cancer by the European Society of Breast Imaging (EUSOBI) and 30 national breast radiology bodies from Austria, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Israel, Lithuania, Moldova, The Netherlands, Norway, Poland, Portugal, Romania, Serbia, Slovakia, Spain, Sweden, Switzerland and Turkey. *Eur Radiol*. 2017;27:2737–43.
- Tabár L, Vitak B, Hsiu-Hsi Chen T, Ming-Fang Yen A, Cohen A, Tot T, et al. Swedish Two-County Trial: Impact of Mammographic Screening on Breast Cancer Mortality during 3 Decades. *Radiology*. 2011;260:658–63.
- Mellado Rodríguez M, Osa Labrador AM. Cribado de cáncer de mama. Estado actual. *Radiología*. 2013;55:305–14.
- Brewer NT, Salz T, Lillie SE. Systematic review: the long-term effects of false-positive mammograms. *Ann Intern Med*. 2007;146:502–10.
- Houssami N, Hunter K. The epidemiology, radiology and biological characteristics of interval breast cancers in population mammography screening. *NPJ Breast Cancer*. 2017;13:3–12.
- Carreira Gómez MC, Estrada Blan MC. Mama densa ¿qué debemos saber? Implicaciones en el cribado. *Radiología*. 2016;58:421–6.
- Karssemeijer N, Otten JD, Roelofs AA, van Woudenberg S, Hendriks JH. Effect of independent multiple reading of mammograms on detection performance. En: Chakraborty DP, Eckstein MP, eds. *SPIE 5372, Medical Imaging 2004: Image Perception, Observer Performance, and Technology Assessment*. Bellingham: Society of Photo-Optical Instrumentation Engineers (SPIE); 2004. p. 82–9.
- Rimmer A. Radiologist shortage leaves patient care at risk, warns royal college. *BMJ*. 2017;359:j4683.
- Wing P, Langelier MH. Workforce shortages in breast imaging: impact on mammography utilization. *AJR Am J Roentgenol*. 2009;192:370–8.
- Skaane P. Breast cancer screening with digital breast tomosynthesis. *Breast Cancer*. 2017;24:32–41.
- Martín SR, Povedano JLR, García MC, Romero ALS, Garriguet MP, Benito MÁ. Prospective study aiming to compare 2D mammography and tomosynthesis+ synthesized mammography in terms of cancer detection and recall. From double reading of 2D mammography to single reading of tomosynthesis. *Eur Radiol*. 2018;28:2484–91.
- Hofvind S, Hovda T, Holen ÅS, Lee CI, Albertsen J, Bjørndal H, et al. Digital Breast Tomosynthesis and Synthetic 2D Mammography versus Digital Mammography: Evaluation in a Population-based Screening Program. *Radiology*. 2018;287:787–94.
- Rodríguez-Ruiz A, Gubern-Merida A, Imhof-Tas M, Lardenoije S, Wanders AJT, Andersson I, et al. One-view digital breast tomosynthesis as a stand-alone modality for breast cancer detection: do we need more? *Eur Radiol*. 2018;28:1938–48.
- Dang PA, Freer PE, Humphrey KL, Halpern EF, Rafferty EA. Addition of tomosynthesis to conventional digital mammography: effect on image interpretation time of screening examinations. *Radiology*. 2014;270:49–56.
- Katzen J, Dodelzon K. A review of computer aided detection in mammography. *Clin Imaging*. 2018;52:305–9.
- Fenton JJ, Taplin SH, Carney PA, Abraham L, Sickles EA, D'Orsi C, et al. Influence of computer-aided detection on performance of screening mammography. *N Engl J Med*. 2007;356:1399–409.
- Nishikawa RM, Bae KT. Importance of Better Human-Computer Interaction in the Era of Deep Learning: Mammography Computer-Aided Diagnosis as a Use Case. *J Am Coll Radiol*. 2018;15:49–52.
- Nishikawa RM, Schmidt RA, Linver MN, Edwards AV, Papaioannou J, Stull MA. Clinically missed cancer: how effectively can radiologists use computer-aided detection? *AJR Am J Roentgenol*. 2012;198:708–16.
- Litjens G, Koo T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A Survey on Deep Learning in Medical Image Analysis. *Med Image Anal*. 2017;42:60–88.
- Dechter R. Learning While Searching in Constraint-Satisfaction Problems. En: AAAI'86: Proceedings of the Fifth AAAI National Conference on Artificial Intelligence. Philadelphia: AAAI Press; 1986. p. 178–83.
- LeCun Y, Bengio Y, Hinton G. Deep Learning. *Nature*. 2015;521:436–44.
- Vaillant R, Monroq C, LeCun Y. Original approach for the localisation of objects in images. *IEE Proc. Vision, Image, and Signal Processing*. 1994;4:245–50.
- Sahiner B, Chan HP, Petrick N, Wei D, Helvie MA, Adler DD, et al. Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images. *IEEE Trans Med Imaging*. 1996;5:598–610.
- Kooi T, Litjens G, Ginneken B, Gubern-Merida A, Sánchez C, Mann R, et al. Large scale deep learning for computer aided detection of mammographic lesions. *Med Image Anal*. 2017;35:303–12.
- Moreira I, Amaral I, Domingues I, Cardoso A, Cardoso M, Cardoso J. INbreast: toward a full-field digital mammographic database. *Acad Radiol*. 2012;19:236–48.
- Heath M, Bowyer K, Kopans D, Kegelmeyer Jr P, Moore R, Chang K, et al. The digital database for screening mammography. Current Status of the Digital Database for Screening Mammography. En: Karssemeijer N, Thijssen M, Hendriks J, van Erning L, editors. *Digital Mammography. Computational Imaging and Vision*, 13. Dordrecht: Springer; 1998.
- Halling-Brown M, Looney P, Patel M, Warren L, Mackenzie A, Young K. The oncology image database (OMI-DB). En: Law MY, CookProc Tessa S, editors. *SPIE 9039, Medical Imaging 2014: PACS and Imaging Informatics: Next Generation and Innovations*, 903906. Bellingham: Society of Photo-Optical Instrumentation Engineers (SPIE); 2014.
- Agarwal R, Diaz O, Lladó X, Martí R. Mass detection in mammograms using pre-trained deep learning models. *J Med Imaging*. 2019;6:031409.
- Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative Adversarial Nets. En: Ghahramani Z, Welling M, Cortés C, Lawrence ND, Weinberger KQ, editors. *Proceedings of the 27th International Conference on Neural Information Processing Systems*. Cambridge (USA): MIT Press; 2014. p. 2672–80.
- Korkinof D, Rijken T, O'Neill M, Yearsley J, Harvey H, Glocker B. High-Resolution Mammogram Synthesis using Progressive Generative Adversarial Networks. 2019, arXiv: 1807.03401.
- Rodríguez-Ruiz A, Lång K, Gubern-Merida A, Broeders M, Genaro G, Clauser P, et al. Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison With 101 Radiologists. *J Natl Cancer Inst*. 2019;111:916–22.
- Liu KL, WuT, Chen PT, Tsai YM, Roth H, Wu MS, et al. Deep learning to distinguish pancreatic cancer tissue from non-cancerous

- pancreatic tissue: a retrospective study with cross-racial external validation. *Lancet Digital Health*. 2020;2:e303–13.
35. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. En: Navab N, Hornegger J, Wells W, Frangi A, editores. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science*. Munich: Springer; 2015. p. 234–41.
 36. Balakrishnan G, Zhao A, Sabuncu MR, Guttag J, Dalca AV. VoxelMorph: A Learning Framework for Deformable Medical Image Registration. *IEEE TMI: Trans Med Imaging*. 2019;38:1788–800.
 37. Trister AD, Buist DS, Lee CI. Will machine learning tip the balance in breast cancer screening? *JAMA Oncol*. 2017;3:1463–4.
 38. Cole EB, Zhang Z, Marques HS, Nishikawa RM, Hendrick RE, Yaffe MJ, et al. Assessing the stand-alone sensitivity of computer-aided detection with cancer cases from the digital mammographic imaging screening trial. *AJR Am J Roentgenol*. 2012;199:W392–401.
 39. Kim E-K, Kim H-E, Han K, Kang BJ, Sohn Y-M, Woo OH, et al. Applying Data-driven Imaging Biomarker in Mammography for Breast Cancer Screening: Preliminary Study. *Sci Rep*. 2018;8:2762.
 40. Becker AS, Marcon M, Ghafoor S, Wurnig MC, Frauenfelder T, Boss A. Deep learning in mammography: diagnostic accuracy of a multipurpose image analysis software in the detection of breast cancer. *Invest Radiol*. 2017;52:434–40.
 41. Geras KJ, Wolfson S, Shen Y, Kim S, Moy L, Cho K. High-resolution breast cancer screening with multi-view deep convolutional neural networks. 2017, arXiv: 170307047.
 42. Rodríguez-Ruiz A, Krupinski E, Mordang J-J, Schilling K, Heywang-Köbrunner SH, Sechopoulos I, et al. Detection of Breast Cancer with Mammography: Effect of an Artificial Intelligence Support System. *Radiology*. 2019;290:305–14.
 43. Salim M, Dembrower K, Eklund M, Lindholm P, Strand F. Range of Radiologist Performance in a Population-based Screening Cohort of 1 Million Digital Mammography Examinations. *Radiology*. 2020;297:33–9.
 44. Kooi T, Karssemeijer N. Classifying symmetrical differences and temporal change for the detection of malignant masses in mammography using deep neural networks. *J Med Imaging*. 2017;4:044501.
 45. Mordang J-J, Janssen T, Bria A, Kooi T, Gubern-Mérida A, Karssemeijer N. Automatic microcalcification detection in multi-vendor mammography using convolutional neural networks. En: Tingberg A, Lång K, Timberg P, editores. *Proc. of the International Workshop on Digital Mammography*. Berlin: Springer-Verlag; 2016. p. 35–42.
 46. Weber RJ, van Bommel RM, Louwman MW, Nederend J, Voogd AC, Jansen FH, et al. Characteristics and prognosis of interval cancers after biennial screen-film or full-field digital screening mammography. *Breast Cancer Res Treat*. 2016;158:471–83.
 47. iCAD. Artificial Intelligence for Digital Breast Tomosynthesis - Reader Study Results (White Paper) 2018. Disponible en: <https://www.icadmed.com/assets/dmm253-reader-studies-results-rev-a.pdf>.(consultada 04.08.20).
 48. Tchou PM, Haygood TM, Atkinson EN, Stephens TW, Davis PL, Arribas EM, et al. Interpretation time of computer-aided detection at screening mammography. *Radiology*. 2010;257:40–6.
 49. Bahl M. Detecting Breast Cancers with Mammography: Will AI Succeed Where Traditional CAD Failed? *Radiology*. 2018;290:315–6.
 50. Kyono T, Gilbert FJ, van der Schaar M. MAMMO: A Deep Learning solution for facilitating radiologist-Machine collaboration in breast cancer diagnosis. arXiv:181102661. 2018.
 51. Rodríguez-Ruiz A, Lång K, Gubern-Mérida A, Teuwen J, Broeders M, Gennaro G, et al. Can we reduce the workload of mammographic screening by automatic identification of normal exams with artificial intelligence? A feasibility study. *Eur Radiol*. 2019;29:4825–32.
 52. Chae EY, Kim HH, Jeong J-W, Chae S-H, Lee S, Choi Y-W. Decrease in interpretation time for both novice and experienced readers using a concurrent computer-aided detection system for digital breast tomosynthesis. *Eur Radiol*. 2019;5:2518–25.
 53. Balleyguier C, Arfi-Rouche J, Levy L, Toubiana PR, Cohen-Scali F, Toledano AY, et al. Improving digital breast tomosynthesis reading time: a pilot multi-reader, multi-case study using concurrent Computer-Aided Detection (CAD). *Eur J Radiol*. 2017;97:83–9.
 54. Benedikt RA, Boatsman JE, Swann CA, Kirkpatrick AD, Toledano AY. Concurrent computer-aided detection improves reading time of digital breast tomosynthesis and maintains interpretation performance in a multireader multicase study. *AJR Am J Roentgenol*. 2018;3:685–94.
 55. Brennan ME, Turner RM, Ciatto S, Marinovich ML, French JR, Macaskill P, et al. Ductal carcinoma in situ at core-needle biopsy: meta-analysis of underestimation and predictors of invasive breast cancer. *Radiology*. 2011;260:119–28.
 56. Shi B, Grimm LJ, Mazurowski MA, Baker JA, Marks JR, King LM, et al. Prediction of occult invasive disease in ductal carcinoma in situ using deep learning features. *J Am Coll Radiol*. 2018;15:527–34.
 57. Lehman CD, Yala A, Schuster T, Dontchos B, Bahl M, Swanson K, et al. Mammographic Breast Density Assessment Using Deep Learning: Clinical Implementation. *Radiology*. 2019;290:52–8.
 58. Kallenberg M, Petersen K, Nielsen M, Ng AY, Pengfei D, Igel C, et al. Unsupervised Deep Learning Applied to Breast Density Segmentation and Mammographic Risk Scoring. *IEEE Trans Med Imaging*. 2016;35:1322–31.
 59. Wanders JO, van Gils CH, Karssemeijer N, Holland K, Kallenberg M, Peeters PH, et al. The combined effect of mammographic texture and density on breast cancer risk: a cohort study. *Breast Cancer Res*. 2018;20:36.
 60. Wang J, Ding H, Bidgoli FA, Zhou B, Iribarren C, Molloy S, et al. Detecting Cardiovascular Disease from Mammograms with Deep Learning. *IEEE Trans Med Imaging*. 2017;36:1172–81.