¿Puede la inteligencia artificial ayudarnos en el diagnóstico y manejo de la osteopo

Categoría: Publicaciones destacadas
Publicado: Miércoles, 19 Julio 2023 13:11
Escrito por: Sofia
Visitas: 4383

Programas de inteligencia artificial son hoy capaces de medir las propiedades de los materiales y las características geométricas de los cuerpos vertebrales, pudiendo predecir el riesgo de fractura.

New Horizons: Artificial Intelligence Tools for Managing Osteoporosis

Nuevos horizontes: herramientas de inteligencia artificial para controlar la osteoporosis

Hans Peter Dimai. The Journal of Clinical Endocrinology & Metabolism, 2023, 108, 775–783

La osteoporosis se define como una enfermedad esquelética sistémica caracterizada por una masa ósea baja y un deterioro de la microarquitectura del tejido óseo con el consiguiente aumento de la fragilidad ósea y la susceptibilidad a las fracturas. Se puede diagnosticar con una DMO igual o inferior a −2,5 desviaciones estándar por debajo del valor medio normal de DMO de adultos jóvenes sanos. Sin embargo, una de las principales limitaciones de este abordaje radica en el hecho de que la mayoría de las fracturas ocurren en puntajes T de â 1,0 a â 2,5 o incluso por encima de â 1,0 (es decir, normal), lo que compromete la sensibilidad de este método *“gold standart”* y su función como herramienta de detección. Se ha demostrado que las fracturas en la columna vertebral, cadera, hombro y muñeca están asociadas con un mayor riesgo de fractura posterior, calidad de vida reducida, discapacidad, y, a excepción de la fractura del antebrazo distal, aumento de la mortalidad.

El término inteligencia artificial (IA) en su sentido actual probablemente se acuñó a mediados de la década de 1950. En la última década se ha logrado un progreso sustancial en muchas áreas de la medicina humana. En general, la IA en medicina se puede dividir en 2 subtipos, a saber, virtual y físico, el primero incluye, por ejemplo, soluciones de imágenes y herramientas de apoyo a la decisión de tratamiento, y el segundo, por ejemplo, prótesis inteligente y cirugía asistida por robot. Con respecto al manejo de la osteoporosis, el subtipo virtual de IA actualmente juega el papel principal, con soluciones disponibles (o en desarrollo) para facilitar el diagnóstico, la evaluación del riesgo de fractura, la detección de fracturas, la evaluación de la calidad ósea y la decisión de tratamiento.

Phoca PDF

Algunos conceptos básicos de la IA en la medicina clínica Muy simplificado y dado que actualmente no existe una definición consensuada internacionalmente, la IA constituye un sistema que combina el poder computacional con conjuntos de datos (idealmente big data) para permitir la resolución de problemas. Una rama típica de la IA es el aprendizaje automático (AA), que utiliza varios algoritmos para aprender de los datos. En la última década, el AA supervisado se ha convertido en el pilar principal de las aplicaciones de atención médica respaldadas por IA. Para entrenar un sistema que se basa en el aprendizaje supervisado, la máquina debe alimentarse con datos que ya están disponibles y son robustos, porque la calidad de estos datos de entrada determinará la calidad de la salida. Por ejemplo, para entrenar una máquina para detectar fracturas de cadera en una radiografía convencional, se debe alimentar el algoritmo de AA con un conjunto de radiografías de cadera convencionales que contengan caderas fracturadas y no fracturadas. Además, hay que decirle al sistema qué caderas están fracturadas y cuáles no, proceso que también se denomina “anotación” de imágenes. En general, cuantas más imágenes anotadas alimenten el algoritmo, mejor será en la detección de fracturas. Las llamadas redes neuronales “convolucionales” (CNN) son las más utilizadas en aplicaciones de atención médica. En general, el desarrollo de un algoritmo de IA específico requiere un conjunto de datos de entrenamiento y un conjunto de datos de prueba. En algunos casos, se reserva un tercer conjunto de datos, sólo con fines de validación. Idealmente, todo el conjunto de datos disponible, como un conjunto de radiografías, se divide aleatoriamente en estos 2 o 3 conjuntos antes del proceso de desarrollo. Esto garantiza que los datos del mismo paciente no se utilicen para más de un conjuntos de datos, lo que comprometería los resultados de las pruebas y la validación, y conduciría a una clasificación y un diagnóstico erróneos en la práctica clínica. El conjunto de datos se utiliza para el entrenamiento y suele ser mucho más grande que el conjunto de datos de prueba, con una relación, por ejemplo, de 80:20 a 60:40.Para encontrar el mejor modelo, el algoritmo desarrollado tiene que ser validado por un proceso cruzado. En la forma clásica de validación cruzada, todo el conjunto de datos se divide en un número específico de grupos de datos (grupos K). En su forma más simple, esto podría ser nuevamente un conjunto de datos de entrenamiento y un conjunto de datos de prueba. En otras palabras, si por ejemplo k es 10, todo el conjunto de datos se dividirá en 10 grupos y se construirán 10 modelos separados. En la primera lectura, la nueve décima parte del conjunto de datos se utilizará con fines de capacitación, mientras que la décima parte de los datos se reservará para las pruebas. En la segunda lectura, se reservará otro grupo para realizar pruebas, mientras que los 9 grupos restantes servirán para fines de capacitación, etc. Este proceso se repetirá 8 veces más. Una vez que se determina que un modelo recientemente desarrollado es suficiente, por ejemplo, para la

Phoca PDF

detección de fracturas de cadera, su desempeño se prueba/valida contra el desempeño humano. En el caso de un algoritmo de detección de fracturas de cadera, podrían ser médicos expertos como radiólogos y/u ortopedistas. Los resultados de rendimiento generalmente se expresan utilizando métricas de rendimiento clásicas, como la sensibilidad, la especificidad y el área bajo la curva (AUC) de la característica operativa del receptor. Si se prueba más de un modelo, se puede elegir el modelo con el mejor rendimiento para realizar más pruebas y validaciones en estudios clínicos, aprobación por parte de las autoridades reguladoras e implementación en la aplicación clínica respectiva.

Diagnóstico de osteoporosis

Las razones para aplicar métodos para el diagnóstico operativo de la osteoporosis son varias. Por ejemplo, el método *gold estándar* para el diagnóstico de la osteoporosis según las recomendaciones de la OMS se basa en la medición de la DMO mediante DXA. Sin embargo, el acceso a los dispositivos DXA depende en gran medida de su disponibilidad. Además de la cantidad de unidades DXA disponibles en un país y algunos obstáculos relacionados con los pacientes, como el tiempo de viaje, las restricciones regulatorias y las barreras para el reembolso, también juegan un papel importante. En la mayoría de los países del mundo, la cantidad de dispositivos DXA disponibles se considera inadecuada. Por lo tanto, existe una clara necesidad de aplicar métodos alternativos que idealmente estarían más disponibles, como la radiografía convencional, y el desarrollo de herramientas compatibles con IA para mejorar el rendimiento diagnóstico de dichos métodos de detección. Por ejemplo, en un estudio preliminar, se aplicó un conjunto de CNN diferentes a las radiografías panorámicas de la boca para diagnosticar la osteoporosis. Dependiendo de la CNN utilizada, los valores de AUC oscilaron entre 0,98 y 0,99. Sin embargo, en lugar de usar DXA como la tecnología de referencia estándar, el diagnóstico de osteoporosis se hizo si se observaron erosiones corticales en las mismas radiografías por dos radiólogos orales y maxilofaciales experimentados. También se han desarrollado softwares basadas en IA para el diagnóstico oportunista de la osteoporosis utilizando radiografías convencionales de la mano y la muñeca. De hecho, en un estudio se utilizó radiogrametría de la diáfisis ósea del tercer metacarpiano y análisis de textura esponjosa del radio distal para entrenar y probar el algoritmo; y DXA como estándar de referencia. El software desarrollado exhibió una precisión cercana al 89%, lo que podría convertirlo en una opción interesante para un uso clínico más amplio. Otros métodos que se han propuesto para el diagnóstico de osteoporosis son, por ejemplo, tomografía computarizada (TC), tomografía computarizada cuantitativa (QCT) y ultrasonido cuantitativo. En general, dichos métodos proporcionan puntajes T que

Phoca PDF

permitirían una categorización en "normal", “osteopenia” y "osteoporosis". Sin embargo, dado que todas estas tecnologías alternativas capturan diferentes propiedades óseas que dan como resultado diferentes gradientes de riesgo, las puntuaciones T obtenidas no pueden intercambiarse simplemente con las puntuaciones T derivadas de DXA. Por otro lado, debe tenerse en cuenta que, independientemente del estado de referencia de la DXA, existe evidencia de que la DMO evaluada mediante QCT de la columna puede mostrar una mayor asociación con el riesgo de fracturas vertebrales incidentes que las puntuaciones T medidas por DXA. Teniendo en cuenta estos aspectos, varios estudios se han esforzado por mejorar la sensibilidad, la especificidad y la precisión de dichos enfoques mediante la aplicación de herramientas de software compatibles con IA. Por ejemplo, en un estudio, se utilizaron tomografías computarizadas de tórax realizadas para la detección del cáncer de pulmón para medir la DMO de la columna vertebral y asignar a los pacientes a una de las categorías de la OMS mediante el uso de un algoritmo de IA totalmente automatizado. El rendimiento diagnóstico resultó ser muy bueno, con un AUC de 0,83 para osteopenia y 0,97 para osteoporosis. Además, se encontró que con cada aumento de 10 UH de los valores de CT, el riesgo de osteopenia disminuyó entre un 32 % y un 44 %, y el riesgo de osteoporosis entre un 61 % y un 80 %. Los autores concluyeron que las TC de tórax de rutina en combinación con IA son de gran valor en la detección oportunista de osteopenia y osteoporosis. Sin embargo, la relevancia clínica limitada de estos hallazgos radica en el hecho de que la medición de la DMO de la columna basada en DXA se realiza utilizando las vértebras lumbares L1–L4, mientras que la medición de la DMO basada en TC de tórax se basa principalmente en las vértebras torácicas. Teniendo en cuenta esta debilidad metodológica, en un estudio se desarrolló un modelo CNN para predecir la DMO a partir de tomografías computarizadas abdominales utilizando DXA de la columna lumbar como estándar de referencia. La osteoporosis se diagnosticó correctamente con un AUC de 0,965 para los conjuntos de datos de validación interna y de 0,970 para los conjuntos de datos externos. Las numerosas herramientas de IA basadas en imágenes para el diagnóstico de la osteoporosis han llevado a una primera revisión sistemática y metanálisis recientemente. Un total de 7 estudios que incluyeron a más de 3000 pacientes fueron elegibles para su inclusión. Utilizando un modelo de efectos aleatorios, la sensibilidad combinada fue de 0,96 y la especificidad combinada fue de 0,95. Sin embargo, como señalaron los autores, los resultados deben interpretarse con cautela debido al alto riesgo de sesgo en la selección de pacientes y la alta heterogeneidad.

Detección de Fracturas Osteoporóticas

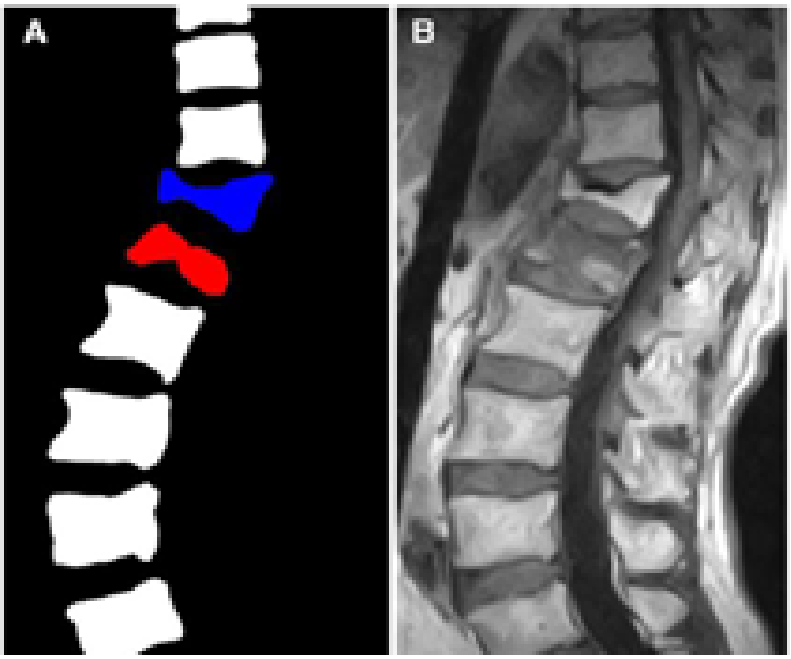
Uno de los pilares en el manejo de la osteoporosis es la detección de

Phoca PDF

fracturas en radiografías convencionales. Por lo tanto, no sorprende que hasta ahora se haya desarrollado una cantidad considerable de herramientas de software de detección de fracturas compatibles con IA. Se ha demostrado que este tipo de herramientas de IA son fiables en el diagnóstico de fracturas y que tienen una alta precisión diagnóstica, similar a la de médicos expertos como radiólogos u ortopedistas. Sin embargo, en un metanálisis se encontró que el rendimiento diagnóstico fue menos convincente cuando se consideraron todos los estudios elegibles para la inclusión, a diferencia de los resultados obtenidos de un análisis de subgrupos que solo incluyó el “grupo de huesos largos” sin vértebras, clavícula y costillas. Este hallazgo es de relevancia clínica ya que proporciona evidencia de que las herramientas de detección de fracturas compatibles con IA pueden funcionar de manera menos confiable si el sitio de interés del esqueleto junto con el tejido circundante, tiene una estructura más compleja. En este sentido, se ha demostrado que muchos de estos problemas de clasificación se deben a los conjuntos de datos de entrenamiento utilizados y a tamaños insuficientes. Las fracturas vertebrales son las fracturas osteoporóticas más abundantes, pero solo un tercio de ellas acudirían a la atención clínica de inmediato, siendo el resto detectadas más o menos por casualidad o en el curso de la evaluación diagnóstica del dolor de espalda crónico. En las radiografías convencionales de la columna vertebral o del tórax, en particular si se realizan por motivos distintos a la exclusión o el diagnóstico de fractura vertebral, se ha demostrado que la tasa de detección es baja. En consecuencia, se han desarrollado programas de concientización como el programa **“Capture the Fracture”** en todo el mundo para mejorar el desempeño diagnóstico de los médicos expertos en este sentido. Además de aumentar la concientización, el soporte técnico en forma de herramientas de software basadas en IA parece ser un complemento lógico. Por lo tanto, la detección automatizada de fracturas vertebrales en radiografías convencionales de tórax y columna se ha convertido recientemente en el foco de atención de los desarrolladores de IA. Por ejemplo, un sistema basado en IA desarrolló un programa de software para la detección de fracturas vertebrales en radiografías laterales de tórax de mujeres ancianas. El software considera una categorización semicuantitativa de estas fracturas según la clasificación de Genant. En general, el rendimiento diagnóstico (es decir, sensibilidad, especificidad y precisión) de esta herramienta parece prometedor, aunque los resultados de rendimiento para las fracturas vertebrales con pérdidas de grado 1 y grado 2 fueron menos convincentes. Dado que las radiografías convencionales de la columna vertebral tomadas inmediatamente después de un trauma a veces no muestran cambios morfológicos, se utilizan además métodos avanzados de imagen, como la TC y la resonancia magnética (RMN). Por ejemplo, la RMN se aplica con frecuencia para detectar edema de la médula ósea como un indicador de la antigÃ¼edad de una fractura vertebral. En este

Phoca PDF

sentido, recientemente se desarrolló un algoritmo basado en IA para la detección automatizada de fracturas vertebrales osteoporóticas recientes, en el que “reciente” se definió como un período de 3 meses después de la lesión respectiva. El algoritmo de IA finalmente elegido fue una combinación de 4 modelos diferentes de CNN, que produjo el mejor rendimiento y fue comparable al de dos cirujanos de columna experimentados. La salida de imagen de esta herramienta de IA proporciona una clasificación codificada por colores de las vértebras en "normal", "fractura reciente" y "fractura antigua" (Figura 1).



A) Imagen después de la detección automática de fracturas y clasificación en "normal" (blanco), "fracturado reciente" (rojo) y " fracturado antiguo (azul)". (B) Imagen original.

Una de las principales limitaciones es que el algoritmo no fue entrenado para identificar fracturas patológicas como las asociadas con la enfermedad ósea metastásica.

El enfoque más reciente y probablemente más avanzado para detectar fracturas vertebrales a partir de TC abdominales y de tórax implica un método de clasificación tridimensional, que no requiere la segmentación de la vértebra individual ni de la TC completa. El método 3D desarrollado muestra un rendimiento excelente con un AUC del 95 % para la detección de fracturas a nivel del paciente y un AUC del 93 % para la detección de fracturas a nivel de vértebras.

Phoca PDF

Riesgo de fractura y predicción de fractura

Para la estimación de la probabilidad de fractura a 10 años, la herramienta más utilizada en todo el mundo es FRAX ®, que es gratuita y está disponible en línea abarcando más de 80 versiones específicas de países y regiones. En los últimos años, se han desarrollado varios modelos basados en IA para la predicción del riesgo de fractura. Por ejemplo, en un reciente estudio prospectivo de cohortes basado en la comunidad, se desarrolló un nuevo modelo de predicción de fracturas y su rendimiento se comparó con el de la versión específica del FRAX ® del país. De los 3 modelos diferentes desarrollados, el que mejor funcionó mostró un AUC de 0,688 para la predicción de fracturas, que fue significativamente mejor que el logrado por FRAX ®. Los principales factores predictivos de riesgo fueron la DMO de la cadera total, la columna lumbar y el cuello femoral. Sorprendentemente, incluso factores como la puntuación subjetiva de artralgia, la creatinina sérica y la homocisteína estuvieron por encima de los predictores convencionales, como la edad o una fractura prevalente. En otro estudio que se basó en datos longitudinales de una cohorte más grande, se desarrollaron modelos basados en CNN utilizando radiografías de columna convencionales. Se encontró que el modelo que sólo usaba radiografías de referencia proporcionaba una predicción del riesgo de fractura vertebral comparable a la del FRAX ®. Utilizando los conjuntos de datos que incluían los resultados de DXA, el rendimiento predictivo del modelo de IA fue incluso mayor que el de FRAX ®. Sin embargo, estos resultados deben interpretarse con cautela, porque FRAX ® proporciona una probabilidad de fractura a 10 años, mientras que ninguno de los estudios mencionados aquí fue diseñado para predicciones que cubren ese período. Además, debe tenerse en cuenta que, independientemente del impresionante rendimiento de estos modelos de IA en la población estudiada, los resultados no pueden extrapolarse simplemente a otras poblaciones, ya que los riesgos de fractura iniciales pueden diferir notablemente entre las diferentes poblaciones.

Propiedades óseas más allá de la DMO

La calidad ósea está determinada no sólo por la DMO, que proporciona información principalmente sobre la cantidad de hueso y su grado de mineralización, sino también por su geometría, microarquitectura y composición tisular. En los últimos años, la estimación no invasiva de la resistencia ósea también se ha vuelto posible *in vivo* (a diferencia del método invasivo de microindentación) mediante el uso de modelos de elementos finitos que se pueden integrar en el software de diferentes modalidades de imágenes como QCT. En un estudio reciente de IA, las propiedades de los materiales y las características geométricas del cuerpo vertebral se extrajeron de imágenes de QCT obtenidas de sujetos

Phoca PDF

masculinos asiáticos, y se desarrolló un algoritmo con el objetivo de proponer un método conveniente y práctico para la predicción clínica de la fuerza de los cuerpos vertebrales. Los resultados del estudio fueron prometedores en términos de capacidad de predicción y consistencia, y los autores concluyeron que el algoritmo desarrollado puede tener un gran potencial para la evaluación no invasiva del riesgo de fractura vertebral. Sin embargo, nuevamente se debe tener en cuenta que los resultados de este estudio no se pueden generalizar a diferentes edades, etnias o sexo femenino. En otro estudio reciente que involucró resultados de DXA y variables clínicas de una población femenina con fracturas prevalentes, se desarrolló un modelo para identificar pacientes propensos a fracturas por fragilidad posteriores. El índice de tensión ósea basado en la CNN desarrollada alcanzó una precisión predictiva cercana al 80 %. con una sensibilidad del 75 % y una especificidad del 84 %. En un estudio basado en tecnología de ultrasonido, se desarrolló un modelo para estimar las propiedades microarquitectónicas del hueso cortical. El modelo final tenía la capacidad de cuantificar la porosidad cortical con gran precisión, lo cual podría ser particularmente útil en el seguimiento de un tratamiento por ejemplo, las propiedades anabólicas óseas de un fármaco.

Apoyo a la decisión de tratamiento

Ha sido un tema de debate en las últimas 3 décadas a quién debemos tratar, cuándo y qué medicamento usar para la osteoporosis, y especialmente para reducir el riesgo de fractura de un paciente. Actualmente, existe un consenso entre la mayoría de las sociedades científicas relacionadas con la osteoporosis de que los pacientes que sufrieron una fractura osteoporótica deben recibir tratamiento con eficacia anti-fractura comprobada. Este enfoque se basa en pruebas sólidas de que una primera fractura, a veces también denominada fractura centinela, es un fuerte predictor de una fractura posterior, con la probabilidad más alta dentro de los primeros 12 a 24 meses después de la fractura. Sin embargo, hay menos consenso sobre cuándo iniciar el tratamiento en un paciente sin fractura prevalente, aunque los conceptos actuales se basan en la inferencia lógica de que si un paciente con una fractura prevalente debe recibir tratamiento para la osteoporosis, cualquier otro debe recibir tratamiento si su probabilidad de fractura es al menos igual a la de un paciente de la misma edad y sexo. Sin embargo, hay diferentes herramientas disponibles para calcular el riesgo de fractura de un paciente, y la categorización del riesgo, por ejemplo, en alto o muy alto, puede diferir dependiendo en los respectivos lineamientos y/o recomendaciones. Dicho esto, una proporción considerable de pacientes con osteoporosis que reciben un tratamiento “adecuado” con eficacia anti-fractura demostrada, no responden. Teniendo en cuenta estos

Phoca PDF

hechos, no sorprende que en el pasado reciente los desarrolladores de software hayan hecho un esfuerzo cada vez mayor para hacer uso de los big datas disponibles en forma de registros médicos electrónicos proporcionados por diferentes sistemas, y proveedores de atención médica. Por ejemplo, un algoritmo de IA desarrollado recientemente para predecir la respuesta de la DMO relacionada con el tratamiento, se basó en los registros médicos electrónicos de más de 15 000 pacientes con osteoporosis seguidos durante un período de 10 años. Además de los 5200 códigos de la Clasificación Internacional de Enfermedades, el algoritmo consideró alrededor de 30 000 resultados de DMO y más de 3500 fármacos diferentes, pero en particular sólo 7 parámetros de laboratorio diferentes, siendo la fosfatasa alcalina total el único de cierta relevancia en lo que respecta al recambio óseo. No se incluyeron la vitamina D ni un marcador establecido de resorción o formación ósea. Sin embargo, de los 7 algoritmos diferentes desarrollados, el que tuvo el mejor rendimiento para predecir la respuesta al tratamiento en términos de aumento de la DMO mostró una precisión de 0,69. Además de cualquier información clínica relevante, una impresión típica de este software basado en IA también proporciona una lista de medicamentos potencialmente elegibles.

> En resumen, en los últimos años se han desarrollado una gran cantidad de algoritmos de IA para facilitar el manejo de la osteoporosis, incluido el diagnóstico, la evaluación del riesgo de fractura, la detección de fracturas, la evaluación de la calidad ósea y la decisión del tratamiento. La comparación de rendimiento entre dichas soluciones y los expertos médicos humanos muestran resultados similares o incluso están a favor del algoritmo de IA. Específicamente, los métodos respaldados por IA para la evaluación de la DMO junto con los datos clínicos, parecen tener un gran potencial para la detección temprana de pacientes que tienen un mayor riesgo de fractura. Sin embargo, incluso si los resultados del estudio parecen prometedores a primera vista, los enfoques metodológicos detrás de un algoritmo de IA recién desarrollado siempre debe leerse cuidadosamente y evaluarse críticamente. El uso de estándares de referencia inadecuados o la selección de variables que tienen poco o ningún valor en la práctica clínica son limitaciones que se encuentran con frecuencia en los estudios de desarrollo de IA. Además, no siempre es evidente que el algoritmo de IA elegido o desarrollado dará como resultado un rendimiento superior al de los métodos estadísticos tradicionales. En consecuencia, existe una clara necesidad de investigación clínica de alta calidad en el campo de la IA aplicado al tratamiento de la osteoporosis. Esto podría lograrse, por ejemplo, estableciendo un marco de mejores prácticas aceptado internacionalmente que tenga en cuenta a los desarrolladores de

Phoca PDF

IA, los expertos en osteoporosis y sus respectivas sociedades científicas, así como a las autoridades sanitarias, incluidas las que participan en los procesos de aprobación.

Phoca PDF