

Premio

Diagnóstico de las lesiones de la calota por tomografía computorizada. Análisis de eficacia por redes neuronales y regresión logística (*)

Estanislao Arana Fernández de Moya (**)

Justificación

Este trabajo de investigación se realiza porque las lesiones de la bóveda craneal o calota sólo habían recibido estudios seriados mediante radiografías simples, realizándose el último en el año 1975 (1). Debido a su sintomatología poco específica y su presentación incidental, los datos con los que se decide una actitud terapéutica o diagnóstica están basados principalmente en la tomografía computorizada (TC) (2). uno de los motivos de este trabajo es conocer las características radiológicas cuantitativas de cada diagnóstico, ya que este tipo de análisis es muy escaso en la bibliografía, apareciendo sólo recientemente artículos en los que se describen para la clasificación de lesiones óseas (3). hasta ahora no se ha comprobado la eficacia de los modelos basados en la regresión logística para la clasificación de las lesiones de la calota, así como tampoco el de las redes

neuronales. Dado que existen pocas referencias en la bibliografía (4) acerca de la comparación de los rendimientos de los distintos modelos de clasificación de las enfermedades; es necesario comparar los obtenidos en este trabajo de investigación con el método idóneo, como son las curvas ROC. Con la construcción de distintos modelos diagnósticos, también se puede estudiar qué variables radiográficas están asociadas a unos u otros diagnósticos y cuáles no son importantes para llegar a esa patología. Se pretende por tanto estimar en qué formas se puede mejorar la eficiencia en el diagnóstico por la imagen de las lesiones de la calota y ver si alguno de estos métodos puede constituir una ayuda al diagnóstico asistido por ordenador.

1. Introducción.

1.1 Epidemiología de lesiones óseas

Las lesiones óseas solitarias suponen un problema que se encuentra con relativa frecuencia en los exámenes radiográficos de cualquier parte del esqueleto. Hasta el momento, en la extensa revisión bibliográfica consultada (desde 1965 hasta ahora), no se ha encontrado un estudio de la frecuencia exacta de la presentación de lesiones óseas focales descritas por cualquier técnica de imagen. En las lesiones óseas focales de los huesos del cráneo se desconoce la frecuencia exacta de su presentación, pero en la época de mayor realización de radiografías simples del cráneo, se consideraba que el siete por ciento mostraban defectos líticos que planteaban diagnósticos distintos a las variantes anatómicas (1). Hasta la fecha no existe un estudio de TC o RM en que se estime la incidencia de lesiones óseas craneales mediante alguna de estas técnicas.

1.2. Orientación habitual

La aproximación habitual inicial a las lesiones óseas solitarias debe descartar que

(*) Premio de la Real Academia de Medicina y Cirugía de Palma de Mallorca 1998

(**) Doctor en Medicina y Cirugía. Valencia

se trate de una variante de la normalidad (5) y en segundo término, distinguir si son benignas o malignas para finalmente intentar clasificarlas según su estirpe histológica. Se debe realizar una aproximación conjunta a todas las lesiones óseas del esqueleto integrando datos de la historia clínica, radiología y anatomía patológica. Esto es así ya que, en frecuentes ocasiones, la información con la que cada profesional contribuye al diagnóstico se complementaría con la que otros miembros podrían aportar, contribuyendo a alcanzar una mayor precisión diagnóstica. En la historia clínica el parámetro más importante es la edad del paciente debido a la acusada distribución de entidades, tumores óseos y cerebrales (6,7). Tras el estudio radiológico, se debe decidir si la lesión es 1) definitivamente benigna y sólo hay que realizar controles radiográficos, o 2) bien es definitivamente maligna o agresiva (dependiendo de la terminología que se use) y por las características del enfermo se sabe cuál es su diagnóstico (p.ej. pacientes con múltiples metástasis) y por ende no es necesario tomar una biopsia; ó 4) finalmente si hasta ese momento el diagnóstico es ambiguo, es necesario realizar una biopsia (7). Existe acuerdo generalizado en realizar el estadiaje de las neoplasias óseas con RM para una mejor delimitación de los cambios locales y extensión a distancia producidos por la lesión.

1.3. Tomografía computarizada

Aun con la información que proporciona la radiografía simple, la TC ha tenido un notable impacto en el proceso de estudio diagnóstico de las enfermedades del sistema esquelético, gracias a su capacidad de proporcionar imágenes exactas de anomalías sutiles, tanto de las estructuras óseas como de los tejidos blandos (8,9). Actualmente se considera el método de elección para el estudio de las lesiones óseas localizadas en sitios donde es difícil la exploración radiológica convencional, como son los huesos planos o la pelvis (10).

La principal ventaja de la TC es que es más precisa que las radiografías de cráneo en la detección de lesiones óseas, dado que la tomografía computarizada puede detectar alteraciones en la mineralización con un mayor grado de sensibilidad que las radiografías simples. La TC puede detectar alteraciones en la mineralización del 15% en los huesos del cráneo, sensiblemente superior al 50% necesario en las radiografías simples para observar un defecto lítico. Además se debe tener en cuenta que lesiones menores de 1,5 cm de diámetro no se visualizan en las radiografías simples y sí en la TC (11). Hay acuerdo generalizado en que la TC es superior a cualquier otro método de diagnóstico por imagen en la evaluación de la calcificación, osificación, adelgazamiento endóstico y las reacciones periósticas, así como para los huesos planos (10). Además de las imágenes con las características de ventana apropiadas para la lesión ósea, se necesitan las adecuadas para el estudio de la partes blandas.

1.4 Eficacia del diagnóstico por la imagen.

La importancia cada vez mayor de los costes de los procedimientos médicos hace que se requieran cada vez más explicaciones acerca de la utilidad de las distintas pruebas diagnósticas. En líneas generales, el objetivo de todo departamento de radiodiagnóstico debería ser producir las mejores imágenes y los diagnósticos más precisos posibles. Pero un análisis más detallado muestra que el diagnóstico por la imagen está involucrado completamente en todo el proceso de la salud del paciente, y que en realidad existen seis niveles en la eficacia del diagnóstico por la imagen (12). Los estudios de eficacia son esenciales a la hora de proporcionarnos una idea para comprender y comparar la utilidad de los exámenes por la imagen.

1.5. Criterios diagnósticos.

1.5.1. Ayudas al diagnóstico médico

Un diagnóstico es una hipótesis acerca de la naturaleza de la enfermedad de una

paciente y que se deriva por inferencia de las observaciones obtenidas de ese sujeto. Cuando se desarrolla el proceso inferencial, nuestra confianza como médicos en un diagnóstico dado se acrecienta reuniendo datos a favor o en contra de la hipótesis inicial. Nuestra tarea no es conseguir la exactitud en el diagnóstico, sino reducir el nivel de incertidumbre diagnóstica lo suficiente para establecer las decisiones terapéuticas más óptimas (13).

Este desmesurado interés en la exactitud diagnóstica ocasiona frecuentemente una excesiva solicitud de pruebas. Se ha comprobado que la confianza en el diagnóstico aumenta cuando se solicitan muchas pruebas complementarias, aunque no haya motivos objetivos para esa mayor confianza, (13). En una época como la actual en la que se busca un mayor ajuste en los costes de todo lo relacionado con la atención sanitaria, el uso razonable de las pruebas diagnósticas pasa a ser de algo deseable a necesario. Nuestros objetivos como médicos son varios; debemos eliminar pruebas diagnósticas innecesarias, preservar las pruebas y procedimientos que producen incrementos sustanciales de los beneficios, comparados con sus coste y riesgos, y evitar restricciones en las pruebas diagnósticas, tan severas, que puedan erosionar la calidad de nuestras decisiones y nuestra confianza en nuestro juicio clínico. Por las citadas razones debemos establecer la exactitud de los medios diagnósticos que disponemos y, una vez conocidos, ver las maneras de poder mejorar nuestra predicción.

Las ayudas al diagnóstico médico se definen como los sistemas de conocimiento activos que usan dos o más características del paciente para generar un consejo específico para ese caso concreto (14). En una definición más general, se entiende como cualquier programa de ordenador o sistema de reglas diseñado para ayudar a los profesionales sanitarios a tomar decisiones clínicas (15).

La necesidad de tales ayudas en el entorno médico es real, dada la complejidad

cada vez mayor del conocimiento médico, el aumento de la fragmentación y especialización de la práctica médica, los costes cada vez mayores de las pruebas diagnósticas y, todo esto unido a la mayor disponibilidad de ordenadores, hace que estos programas sean fácilmente accesibles y utilizables (16).

Estas necesidades se han hecho más visibles en el ámbito del diagnóstico por la imagen, ya que además de las complejidades anteriores que conlleva cualquier decisión médica, se pueden añadir las inherentes al diagnóstico por la imagen, que son: la inadecuada interpretación de la imagen porque existe ruido o es visualmente complicada, su relación con la anatomía es compleja, las lesiones son sutiles, la relevancia clínica de la lesión depende de una gran cantidad de características y la variedad de diagnósticos diferenciales es muy grande (17). Todos estos posibles problemas han suscitado un gran número de ayudas al diagnóstico por la imagen, que han sido evaluadas desde distintos puntos de vista mediante técnicas estadísticas e informáticas (18,19). En esta búsqueda se plantean trabajos como el presente para lograr al menos estimar el conjunto mínimo de variables necesario para un diagnóstico radiológico preciso.

1.5.2. Redes neuronales

La mayoría de fenómenos naturales, entre ellos los humanos, con el curso de los años se ha visto que no pueden ser explicados en base a los sistemas lineales tradicionales newtonianos. Los fenómenos biológicos se caracterizan por ser dinámicamente complejos, tener una conducta difícilmente previsible y con un caos matemático, lo que lleva a que la mayoría de investigaciones, desde varios campos de la ciencia, hayan concluido que estos fenómenos sólo pueden explicarse mediante funciones no lineales (20).

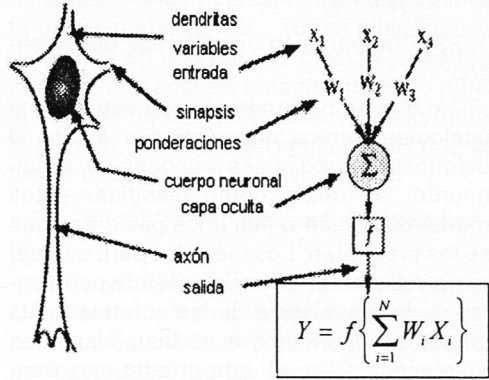
Como se ha señalado previamente los fenómenos humanos presentan una gran complejidad, por lo tanto muchos detalles se nos pueden escapar si se analizan de

forma lineal. Hasta la fecha, la mayoría de los análisis de los procesos fisiopatológicos se han basado en metodología estadística lineal, y las técnicas no lineales no están, hasta el momento, muy estandarizadas, aunque cada vez cobran más importancia. La revisión hasta la fecha de las publicaciones de estadística clínica que llevan a herramientas de decisión, muestra que una gran cantidad de ellas no obtienen una alta precisión (20). Casi la totalidad de estas reglas estadísticas predictivas se basan en los métodos multivariantes, teniendo estas reglas sólo un moderado éxito y poca transportabilidad a otros entornos clínicos. Estos modelos lineales se basan en la invariabilidad de sus constantes. Si una significativa porción de la fisiología y patología humana es compleja y manifiesta una conducta caótica, estas constantes pueden volverse muy variables rompiendo así cualquier modelo lineal.

Una aproximación a la predicción clínica, que se plantea con un alto grado de precisión, es el análisis discriminante cuadrático (21). el hecho de que este método no lineal fuera más preciso, comparado con otros métodos lineales, apoyó la hipótesis de que el uso del análisis no lineal puede mejorar apreciablemente la exactitud diagnóstica de la decisión médica. Desafortunadamente, esta difícil aproximación estadística no ha encontrado una amplia aceptación (20).

Un reciente acercamiento al mundo de la clasificación y el diagnóstico no lineal proviene de las redes neuronales. Las redes neuronales, en una visión general, son una manera de procesamiento de la información mediante un programa de ordenador. En ellas se usan patrones matemáticos de reconocimiento para "aprender" las complejas interacciones entre las variables de entrada (independientes) y las de salida o dependientes. Posteriormente se usan los patrones aprendidos para estimar la salida cuando se le presenten nuevos pacientes. Así la red neuronal adquiere "conocimiento" basado en un aprendizaje previo, algo similar cuando en medicina utilizamos el térmi-

no juicio clínico para una decisión basada en nuestro aprendizaje previo y experiencia (22,23). Las redes se disponen básicamente en capas. La capa de entrada es donde se registran las variables independientes o predictoras, la capa oculta es donde se localizan las unidades de procesamiento no lineales y la capa de salida es donde se expresa la variable dependiente. En una analogía con los sistemas biológicos se podría comparar mediante el siguiente esquema:



Comparación de neurona biológica y artificial

Las redes neuronales se desarrollan como una forma de inteligencia artificial similar a un modelo de las funciones neurológicas humanas. El "entrenamiento" de las redes neuronales es el proceso por el que las ponderaciones, o ajustes entre unidades de proceso, se determinan para cada conexión, que al igual que las sinapsis pueden ser excitatorias e inhibitorias. Cada vez que se presenta el caso a la red neuronal es una iteración y en cada iteración las ponderaciones se ajustan para obtener la salida más acorde con la realidad. Tras un determinado número de iteraciones se acepta que la red neuronal ha preparado todos sus ponderaciones para toda clase de nuevas variables que le lleguen.

Los psicólogos cognitivos que han estudiado las maneras en que se formula el diagnóstico médico, han demostrado que la estructura del razonamiento médico cambia

con el tiempo. Cuanto más avezado sea el médico, su juicio clínico dependerá de la interpretación de patrones basados en su experiencia personal con casos vividos, más que en principios científicos, reglas aprendidas o descripciones de libros de texto (24). A la vista de estas interpretaciones del conocimiento médico, el entrenamiento de las redes neuronales simula el desarrollo de la experiencia en médicos expertos, p. ej. patrones aprendidos por exposición secuencial a casos, más que reglas aprendidas, basadas solamente en modelos de patogénesis u opiniones expertas (24).

1.5.3. Comparación de sistemas diagnósticos

Lo que se pretende cuando estudiamos cualquier técnica diagnóstica o ayuda al diagnóstico médico es conocer su rendimiento, es decir, cómo clasificar estos modelos de bien o mal a los pacientes que se les presentan. Los métodos para evaluar sus rendimientos deberían ser independientes de la prevalencia de las enfermedades que se estudian, ya que podrían falsear los rendimientos (p. ej. una prueba muy sensible en una población con alta prevalencia de enfermedad podría dar un rendimiento altísimo) (25). El método ideal debe ser, por tanto, aquel que sea independiente de las frecuencias de enfermos y sanos y que tampoco dependa al decidir del sesgo del sistema, esto es, la tendencia particular de escoger una alternativa sobre otra (estar enfermo frente a no enfermo o viceversa). Para mostrar la precisión de un modelo que clasifica en una categoría dicotómica (p. ej. cáncer o no cáncer), existen varias formas de comparar las maneras en que lo hace cada uno, siendo las curvas ROC (receiver operating characteristics, característica operativa del receptor), las preferidas en la actualidad (17,25).

2. Objetivos

Una vez conocidos a través de la introducción los problemas que conlleva el estudio de las lesiones de la bóveda craneal, a

continuación se ponen de manifiesto los objetivos de este trabajo de investigación.

- Estimar cuál es la frecuencia real de las lesiones de calota observadas mediante tomografía computarizada en un hospital de referencia.

- Mostrar cuál es la distribución de las lesiones histológicas encontradas entre estas entidades que afectan a la calota.

- Describir las características clínicas y radiológicas fundamentales con las que se presentan las distintas histologías halladas.

- Construir modelos, tanto de regresión logística como de redes neuronales, para el diagnóstico y clasificación de las lesiones halladas en la calota, a través de las variables clínicas y radiológicas.

- Comprobar la eficacia diagnóstica de estos modelos de clasificación mediante el análisis de las curvas ROC.

- Describir las variables seleccionadas por el modelo estadístico y de red neuronal para cada problema diagnóstico.

- Determinar qué modelo, estadístico o de red neuronal, es el más eficaz para la ayuda al diagnóstico de las lesiones de la calota.

3. Pacientes

3.1. Muestra de estudio

Es un estudio de evaluación de pruebas diagnósticas, realizado de forma prospectiva, en el que se analizaron las imágenes de la TC. Seleccionamos todos aquellos pacientes registrados en un hospital universitario de referencia, que cumplieran los criterios de inclusión, desde Enero de 1992 hasta Diciembre de 1996. La muestra incluye tanto pacientes pediátricos como adultos con lesiones en la calota descubiertas de manera sintomática o incidental. Para su comprobación se ha requerido de citología o histología. En aquellos casos en los que alguna de estas técnicas patológicas no se realizó -un total de 22 pacientes-, se consideró necesario un seguimiento tanto clínico como radiológico de al menos tres años, en

este seguimiento se ha tenido en cuenta que el diagnóstico inicial y los posteriores controles tanto clínicos como radiológicos fueran concordantes con ese diagnóstico.

3.2. Criterios de inclusión

El criterio principal de inclusión fue toda lesión focal situada en la calota como única localización conocida al momento de realizar la tomografía computerizada (TC). Las lesiones fueron evaluadas a la primera presentación del paciente por su lesión de calota, aunque estuvieran disponibles todas las exploraciones de imagen de la evolución de la patología craneal.

3.3 Criterios de exclusión

Se excluyeron del ámbito de estudio de este trabajo de investigación las variantes anatómicas y las lesiones congénitas, traumáticas o yatrogénicas, puesto que normalmente éstas no son incluidas en el diagnóstico diferencial de las lesiones focales de calota.

4. Métodos

4.1. Tomografía computerizada

Entre todas las técnicas de diagnóstico por la imagen existentes, la tomografía computerizada se ha considerado técnica básica para el estudio por varias razones, una de ellas es su fácil disponibilidad y el ser una técnica que evalúa los cambios óseos corticales mejor que las radiografías simples y la RM. Para llevar a cabo este estudio ha sido imprescindible que las imágenes de TC estuvieran impresas con una amplitud y nivel de ventana adecuados. Se han utilizado dos equipos de TC para recoger las imágenes:

- En el equipo General Electric CT 9800, las imágenes se vieron con unas características de imagen $wL 250 \pm 55$ UH, $wW 1500 \pm 125$ UH y un campo de visión - field of view (FOV)- de 18-24 cm.

- Los pacientes explorados mediante el General Electric CT Pace se visualizaron

con los parámetros $wL340 \pm 58$ UH, $wW1480 \pm 165$ UH con un FOV de 18-24 cm.

En todas las exploraciones se utilizó un grosor de corte entre 2 y 5 mm con intervalos entre los cortes entre 3 y 10 mm.

4.2. Variables

Las imágenes de cada caso fueron individualmente revidadas por dos de los autores sin conocimiento del diagnóstico histológico. Diecinueve características de la imagen así como datos anatómicos y demográficos se evaluaron sin conocimiento del diagnóstico final. Todas ellas fueron recogidas en una hoja de cálculo y usadas tanto para el análisis estadístico como para la red neuronal. Cuando una variable tenía dos o más hallazgos en la misma lesión, la forma más grave fue la registrada.

4.2.1. Variables radiológicas independientes

Se consideraron como tales a localización, tamaño, forma, número de lesiones, patrones lesionales, expansividad, lobulación de los bordes, borde escleroso, reacción perióstica, matriz, masa de partes blandas.

4.2.2. Variables no radiológicas independientes.

Se establecieron variables provenientes de la historia clínica tales como la edad, sexo del paciente, síntomas y duración de los mismos.

4.2.3. Variables dependientes

Se consideraron variables dependientes para el análisis estadístico y mediante red neuronal a benigno/maligno y cada uno de los diagnósticos histológicos, siendo estos: histiocitosis de células de Langerhans, quiste epidermoide, osteoma, meningioma, hemangioma y displasia fibrosa. Los restantes diagnósticos histológicos, debido a su baja frecuencia, se agruparon en otros benigno y otros maligno. Estos dos grupos, por su heterogeneidad, se utilizaron para el análisis estadístico y mediante red neuronal sólo para benigno y maligno, quedando excluidos del resto de procesos diagnósticos.

Aunque el meningioma frecuentemente presenta benignidad histológica, se decidió incluirlo como maligno por diversos motivos. Aunque la mayoría de los meningiomas encontrados en series de autopsias son pequeños y asintomáticos, los que ocasionan cualquier tipo de síntomas son de mayor tamaño (26). Estos tumores, aunque raramente presentan características malignas (27), pueden dar metástasis y elevación del antígeno carcinoembrionario sin presentar características histológicas especiales (28,29). Además presentan una alta frecuencia de recurrencias, incluso en aquellos casos que son más fácilmente resecables (30). Por tanto, una de nuestras razones es que, debido a su especial sintomatología y a su localización, se tiende a tomar una conducta intervencionista frente a la mayoría de los meningiomas por su agresividad clínica en aquellos casos sintomáticos (30,31). Además, en los estudios donde un tumor presenta una conducta biológica ambivalente, con benignidad histológica pero con malignidad clínica, se clasifican como malignos (32,33).

Se decidió incluir los pacientes con mieloma en la variable metástasis ya que su comportamiento clínico y radiológico es muy semejante al de éstas últimas, aunque sean tumores óseos primarios (34). Así mismo los pacientes con quistes leptomeníngeos, aunque etiológicamente son debidos a fracturas, su presentación clínica puede ser muy retrasada respecto al episodio traumático, no presentando ninguna línea de fractura, pudiendo ser una lesión-radiolúcida solitaria o no recordando el paciente ese episodio; por estos motivos se ha decidido incluirlas en el estudio codificándolas en la categoría de lesiones benignas (35).

4.3. Análisis estadístico mediante regresión logística

Se contruyeron una ecuación de regresión logística para malignidad y siete modelos para los diagnósticos más frecuentes. el método para realizar la ecuación fue el

stepwise forward tomando sólo como variables significativas aquellas con $p < 0.05$. La magnitud del efecto se calculó con el uso de odds ratio ajustados y un intervalo de confianza del 95% (IC).

Se prefirió la regresión logística al análisis discriminante por varias razones: no todas las variables independientes eran normales y habían tanto cualitativas como cuantitativas requiere menos aproximaciones teóricas, en la práctica es más robusta estadísticamente, más adecuada para muestra pequeñas y es más fácil de comprender que el análisis discriminante (36). la única desventaja es que la RL tuvo varias ecuaciones para los distintos diagnósticos.

4.4. Red neuronal

Las redes neuronales (RN) suponen una forma de procesamiento no lineal de la información, que intenta simular la manera en la que se procesa ésta en los sistemas biológicos (37). las variables independientes suponen los nodos de entrada, éstos se conectan, por unas ponderaciones a las neuronas ocultas, que son las que procesan la información, y éstas se conectan con los nodos que ofrecen la salida o respuesta. Estas neuronas se disponen en capas, estando por tanto constituidas una red neuronal por una capa de entrada, al menos una capa oculta y una capa de salida. El patrón básico de las redes neuronales es que aprenden de sus errores. Cada vez que se introducen las variables independientes y se obtiene una respuesta se conoce como iteración, una vez conocida esta respuesta se compara con la real y se ajustan los pesos matemáticos de forma retrógrada. el ajuste de los pesos se realiza hacia detrás con cada iteración y las redes neuronales que realizan este proceso, se las conoce como backpropagation o propagación retrógrada (37).

Simultáneamente se diseñaron dos modelos de redes neuronales que permitieran el diagnóstico de benignidad y malignidad y de las siete entidades más frecuentes encontradas, con las mismas variables utilizadas para el análisis estadístico.

4.4.2. Estimación de rendimientos

La comparación de los resultados de cada una de las distintas redes neuronales se puede realizar por distintas técnicas: la media del error al cuadrado, el error medio o el área bajo la curva ROC de cada una de las redes (18,38). Se han utilizado en el presente estudio las curvas ROC por dos razones: muestran una estimación de la probabilidad máxima basada en una salida continua, como la producida por la red neuronal y permiten la comparación estadística con otras curvas ROC obtenidas con otras redes neuronales, otras técnicas diagnósticas o métodos diagnósticos como los presentes de regresión logística (18).

4.4.3. Función de activación neuronal

La función de activación asocia la entrada a la neurona con la salida, $y = F(x)$, siendo y la salida de la neurona y x la entrada. También existen gran diversidad de funciones de activación, aunque principalmente nos encontraremos con tres: función umbral, función lineal y función sigmoide.

Las redes implementadas utilizan como función de activación la sigmoide bipolar,

$$\text{SALIDA} = f(x) = \frac{2}{(1+e^{-x})} - 1,$$

resultando que los valores de salida están comprendidos entre -1 y 1. Siendo su derivada

$$f'(x) = \frac{1 - \text{SALIDA}^2}{2}$$

Finalmente queda normalizada la salida entre 0 y 1 constituyendo la probabilidad de que el paciente obtenga una correcta clasificación para cada uno de los diagnósticos.

4.4.4. Diseño y optimización de la red neuronal

La construcción de una red neuronal desafortunadamente no está ajustada a ningún modelo teórico de una red neuronal ideal (39), debiendo realizarse en forma de ensayo y error, siendo imprescindibles en su construcción una validación interna y otra externa (37). La validación interna

consiste en que el diseño de la misma sea lo mejor posible, con los parámetros y los casos de entrenamiento disponibles. La validación externa se trata de que una red con la arquitectura adecuada y el suficiente entrenamiento sea capaz de generalizar, esto es, dar respuestas correctas para casos que no ha visto durante su fase de aprendizaje (p.Ej. diagnósticos correctos en futuros pacientes)(18).

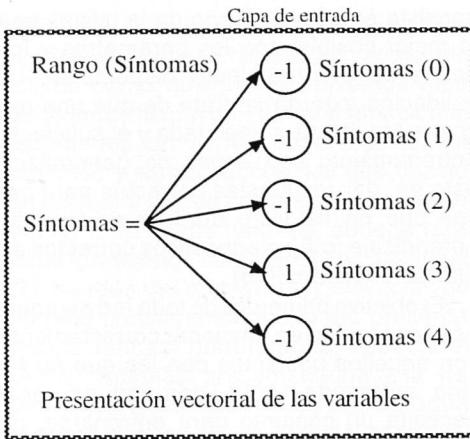
El objetivo primordial de toda red neuronal es la capacidad de funcionar correctamente con aquellos pacientes con los que no ha sido entrenada (40). Toda red neuronal necesita un conjunto para entrenarse, un conjunto de pacientes de prueba para ver cual es la mejor entre los diferentes diseños posibles y un conjunto de validación, que son los casos futuros que se puedan presentar en los que es necesario la ayuda de la red neuronal para su predicción.

La validación cruzada es descomponer la muestra en diversas porciones y entrenar a la red neuronal con una parte y probarla con la otra; p. ej. dividirla por la mitad y entrenar con una mitad y probarla con la otra y viceversa. Un mayor refinamiento de esta técnica es el método leave-one-out (dejar uno fuera), ya citado previamente para la estimación de ambos modelos.

4.4.5. Red neuronal empleada

La red neuronal utilizó como variables de entrada las mismas utilizadas anteriormente en el análisis estadístico siendo en total 19 las consideradas (Ilustración 1).

Las variables de entrada a la red se normalizan entre -1, que expresa reposo y 1, que expresa activación. Para introducir a la red las variables de tipo discreto, como síntomas del paciente, se hace corresponder a cada una con tantas neuronas en la capa de entrada como rango de la variable, tal y como ilustra la siguiente figura; activando a 1 la neurona correspondiente al valor de la variable y a -1 el resto. De esta forma se obtienen a la entrada de la red vectores ortogonales para cada variable discreta de entrada, cuya representación escrita para el ejemplo del gráfico sería síntomas (-1,-1,-1,1,-1)



Al hacer esto obtenemos un total de 63 neuronas en la capa de entrada. Con este sistema están correctamente representadas todas las distintas características de las variables discretas en el espacio dimensional. Este método es mejor que representar los distintos síntomas en una escala de 0 a 1, porque por ejemplo síntomas 0 es 0, síntoma 1 =0.33 síntoma 2 =0.67 y así sucesivamente. Esta aproximación es más sencilla matemáticamente pero potencialmente errónea, ya que el síntoma 2 no tiene que ser necesariamente el doble de importante que el síntoma 1 para la red neuronal.

La salida de las redes es una variable continua que expresa la probabilidad de que el paciente estudiado tenga cada uno de los diagnósticos buscados. Matemáticamente se ha comprobado que estas probabilidades asignadas por la red neuronal corresponden a las probabilidades a posteriori del análisis bayesiano (41).

En el primer estudio la red sólo aprenderá a clasificar a los pacientes en benignos o malignos, dando una probabilidad para benignidad o bien para malignidad; mientras que la segunda red neuronal realiza una clasificación más detallada del paciente, en uno de los siete grupos diagnósticos. En esta red para los diagnósticos a cada paciente se le asigna una probabilidad de pertenecer a uno de los grupos diagnósticos por lo tanto para cada paciente se obtienen siete probabilidades. Se considera

verdadero positivo cuando el número más alto es el asignado al diagnóstico al cual realmente pertenece y falso negativo o incorrectamente clasificado cuando se le asigna un número más alto a alguna de las clases distintas a las cuales el paciente pertenece. Los restantes pacientes, incluidos en los grupos otros benignos y otros malignos, sólo son clasificados en la red neuronal utilizada para benigno y maligno.

4.5. Curvas ROC: Comparación entre regresión logística y red neuronal.

Clásicamente, la exactitud de una prueba diagnóstica se ha evaluado en función de la sensibilidad y especificidad. No obstante, éstas varían en función del criterio elegido como punto de corte entre la población sana y la enferma y de la prevalencia de la enfermedad entre la muestra estudiada. Una forma más global de conocer la calidad de la prueba en el espectro completo de puntos de corte, es mediante el uso de curvas ROC (receiver operating characteristics, características operativas del receptor, también llamadas curvas de rendimiento diagnóstico).

La curva ROC es un gráfico en el que se observan todos los pares sensibilidad/especificidad resultantes de la variación continua de los puntos de corte en todo el rango de resultados observados (42). Estas curvas se idearon en el campo del radar como una forma de mostrar las proporciones entre la señal y el ruido. Si se piensa que las positivas verdaderas son la "señal correcta" de una prueba de diagnóstico y que las positivas falsas son el "ruido", es posible ver la forma en que este concepto se aplica.

En el eje Y de coordenadas se sitúa la sensibilidad o fracción de verdaderos positivos (FVP) calculada en el grupo de diagnósticos positivos para la categoría estudiada. En el eje X se sitúa la fracción de falsos positivos (FFP) ó 1-especificidad, definitiva como $FP/VN+FP$ y calculada en el subgrupo no afectado. Cada punto de la

curva representa un par S/1-E correspondiente a un nivel de decisión determinado. Una prueba con discriminación perfecta, sin solapamiento de resultados entre las dos poblaciones, tiene una curva ROC que pasa por la esquina superior izquierda, donde S y E toman valores máximos (S y E=1). Una prueba sin discriminación, con igual distribución de resultados en los dos subgrupos, da lugar a una línea diagonal de 45°, desde la esquina inferior izquierda hasta la superior derecha.

El área bajo la curva (A_z) es una medida global de la exactitud de una prueba diagnóstica. Se define como la probabilidad de clasificar correctamente un par de individuos sano y enfermo (enfermo correspondiente al diagnóstico estudiado y sano a no tener ese diagnóstico), seleccionados al azar entre la población, mediante los resultados obtenidos al aplicarles la prueba diagnóstica. Es decir, es la probabilidad de que el resultado de la prueba resulte más anormal en el paciente enfermo $A_z \text{ ROC} = \text{Prob}(X_{\text{enfermo}} > X_{\text{sano}})$. Por convenio, el $A_z \text{ ROC}$ es siempre mayor o igual que 0.5. La capacidad de discriminación de la prueba diagnóstica puede evaluarse estimando el intervalo de confianza del $A_z \text{ ROC}$. Si el intervalo no incluye el valor 0.5 la prueba es capaz de discernir entre enfermos y sanos. Las curvas ROC tiene dos usos primarios. Uno, para comparar dos tipos de clasificación; si el área bajo la curva ROC para el método A es mayor que para la curva B, entoces se concluye que A es la mejor opción. Esto es debido a que en cualquier valor de verdadero positivo (el mismo para ambas clasificaciones), A tendrá un menor valor de falsos positivos que B; de forma similar ante un determinado valor de falsos positivos, A tendrá el mayor de los verdaderos positivos. Evidentemente la prueba A es preferible a B. El segundo uso es para escoger el nivel que lleva a la clasificación más adecuada de sensibilidad y especificidad, según la prevalencia de la enfermedad y la importancia relativa de los falsos positivos en un lado y los falsos negativos en el otro (43).

Para aportar más información al resultado de la curva ROC se incluyen los resultados de sensibilidad y especificidad en el punto de máxima eficacia diagnóstica. Debido a que es la información de un sólo punto, este dato debe observarse siempre en conjunción con el área bajo la curva, que es el estimador más fiable del rendimiento del modelo diagnóstico estudiado. Con el fin de comparar la capacidad predictiva de los modelos logísticos y de redes neuronales, se construyeron las curvas ROC correspondientes a cada modelo, en base a las probabilidades predichas para cada individuo. Las probabilidades predichas tanto por la RN como por la RL fueron comparadas mediante áreas bajo la curva ROC según un método no paramétrico de dos colas (44).

4.6. Programas informáticos

Se emplearon los programas SPSS 6.1.2. (SPSS Inc) con licencia universitaria y EPIINFO 6.04, distribuido gratuitamente por el CDC de Atlanta y la OMS. Este último se utilizó para realizar las tablas de 2X2, calcular la prueba exacta de Fisher, *odds-ratio* y sus intervalos de confianza. Siendo realizado el *leave-one-out* mediante el SPSS y la hoja de cálculo Excel 7.0 (Microsoft Corporation). La red neuronal fue programada en lenguaje C+ con su propia rutina para desarrollar el *leave-one-out*. Para el cálculo y comparación de las curvas ROC hemos utilizados dos programas LABROC1 Y CLABROC, distribuidos gratuitamente por Charles E. Metz, Departamento de Radiología, Universidad de Chicago, EE.UU (45).

5. Resultados

5.1. Muestra de estudio

De los 4.012 TC craneales realizadas durante el periodo de estudio, se presentaron 359 lesiones de la calota distintas a las variaciones anatómicas. Con los criterios definidos previamente en 167 pacientes, se obtuvo una comprobación clínica o

histológica, lo que resulta en un 4.16% de lesiones comprobadas de calota y en una incidencia del 8.94% de hallazgos en la bóveda craneal, no concordantes con variaciones anatómicas. Confirmación quirúrgica se obtuvo en 145 casos. En 22 pacientes el diagnóstico se estableció por seguimiento clínico o radiológico durante al menos tres años.

La muestra objeto de estudio comprende por tanto 167 pacientes, 74 hombres (44,6%) y 93 mujeres (55,4%), con una medida de edad de $34,11 \pm 25,52$ años, rango: 6 meses a 81 años. El número total de lesiones benignas fue de 122 (73,1%), con un rango desde 5 a 80 años ($26,1 \pm 22,9$); y 45 malignas (27,0%) con un rango de edad desde 6 meses a 84 años ($55,7 \pm 18,6$).

Los síntomas más comunes de inicio fueron la tumoración palpable (74,7%), síntomas neurológicos (15,9%), cefalea (5,3%) y dolor local (4,2%). Las duraciones de los mismos comprendía un rango desde un mes hasta 26 años. En 18 pacientes (10,8%) las lesiones de la calota fueron un hallazgo incidental durante la realización de la TC.

Las lesiones más comunes se describen en la Tabla 2. Del total de pacientes, 145 de los mismos presentaron una lesión solitaria y en 22 hubo más de una lesión, que se localizaron predominantemente en el hueso parietal (75, 44,9%) seguido del frontal (63, 37,7%), occipital (26, 15,6%) y de forma más rara en suturas o fontanelas (3, 1,8%). Su forma fue más frecuentemente ovoide (96, 57,5%) que circular (71,42,5%). Las lesiones mostraron en su mayoría un borde bien diferenciado (78,46,7%), siendo la forma más infrecuente el borde pobremente definido (27, 16,2%). El patrón de las mismas fue, de forma más habitual, la apariencia lítica geográfica (70, 41,9%), presentándose la forma mixta permeativa (6, 3,6%) como menos común.

Los bordes de las lesiones no presentaron esclerosis en la mayoría de los pacientes (88, 52,7%), seguido de los casos con bordes esclerosos (38, 22,8%); ésta característica no fue aplicable en el resto de

pacientes (41,24,6%). Esto es debido a que la alta atenuación de la lesión hace que no se puedan diferenciar sus bordes lesionales. Del total de la muestra, los bordes fueron más comúnmente lobulados (121,72,5%) que lisos en su presentación radiográfica.

El centro de las lesiones estuvo en el diploe de forma habitual (87,52,1%), seguido por las centradas en la tabla externa (32,19,2%), siendo la presentación menos común en la tabla interna (3, 1,8%). La afectación ósea fue la manera más frecuente en ambas corticales (121, 72,5%), seguida de la cortical externa (27, 16,2%), produciéndose únicamente lesiones diploicas en dos pacientes 91,2%). La forma en que se produjo la afectación de las corticales fue con rotura de las mismas en la mayoría de pacientes (125,74,9%), seguido de su engrosamiento (29 17,4%) y finalmente el adelgazamiento de las mismas (11, 6,6%). El hueso de la calota craneal afectado por una lesión presentó expansividad en grado moderado como forma más habitual (75, 44,9%), seguido de la afectación como formas leves (47, 28,1%), en tercer lugar las apariencias severas (40, 24%), y no existiendo expansividad alguna en cinco pacientes (3%), es decir, eran puramente diploicas. Esta afectación ósea se acompañó de masa de partes blandas de presentación subgaleal (57, 34,1%), seguido de masa a ambos lados de los huesos craneales (53,31,7%) y como forma menos prevalente, la intracraneal (15, 9%). Destaca la ausencia de este componente de partes blandas en 42 pacientes (25,1%).

Respecto a la matriz de las lesiones, la forma más habitual de presentación fue sin la misma (75, 44,9%), seguido del componente calcificado (24, 14,4%), siendo las menos frecuentes el vidrio esmerilado y el secuestro (8, 4,8% cada una). Generalmente no se acompañaron de reacciones periósticas (151, 16%), apareciendo éstas sólo en 16 pacientes (9,6%). En aquellos casos con contraste intravenoso en la TC (96, 57,4%), se observó realce de las lesiones en 58 pacientes y en 38 de ellos no se observaron cambios.

5.2. Regresión Logística

Las variables incluidas en la ecuación de RL fueron edad, definición de bordes, masa de partes blandas y número de lesiones (Tabla 4). La variable número de lesiones entró en el modelo de regresión logística stepwise porque mejoró significativamente la verosimilitud del modelo, aunque según el test de Wald presenta un valor de p al límite de la significación estadística ($p=0,053$). Dado que para la selección de variables prevalece el estadístico de razón de verosimilitudes sobre el test de Wald, se decidió mantener dicha variable en la ecuación final.

Las variables que importaron para el diagnóstico histológico están resumidas en la Tabla 5. En la histiocitosis de células de Langerhans las variables incluidas en dicha ecuación fueron, lobulación, forma y edad. Osteoma tuvo como variables importantes centro, edad, diámetro y duración de los síntomas. Para quiste epidermoide la única variable seleccionada fue el hueso. En las metástasis las variables escogidas fue definición de bordes, edad y número de lesiones. El sexo del paciente y la reacción perióstica contaron para el meningioma. El diagnóstico de hemangioma se basó en la forma, masa de partes blandas y reacción perióstica. La displasia fibrosa dependió del diámetro y edad seleccionadas en la ecuación de regresión logística.

5.3. Red neuronal

En la RN para malignidad las variables fueron encabezadas por edad y definición de bordes (Tabla 6). La exclusión de las variables desde abajo hacia arriba, proporcionó $A(z)$ similares hasta la exclusión de la definición de los bordes. Cuando la ésta fue excluida, se produjo un descenso significativo de $A(z)$ en la RN con edad como única variable ($0,8131 \pm 0,0331$ vs $0,9129 \pm 0,0031$, $p < 0,05$). Por tanto, las variables mínimas necesarias para un diagnóstico con confianza de malignidad y benignidad son edad y definición de los bordes. En la RN para diagnósticos histológicos el orden de las

variables se muestra en la Tabla 6. Las diferentes $A(z)$ en la RN para diagnósticos y cada enfermedad son mostradas en la Tabla 7. La eliminación de las variables no disminuyó el $A(z)$ hasta que la edad fue eliminada. Entonces la RN mostró un descenso de la $A(z)$ para: histiocitosis de células de Langerhans ($p < 0,001$), meningioma ($p < 0,001$) y osteoma ($p < 0,05$) con una pérdida no significativa del área bajo la curva para los restantes diagnósticos. Esto significa que un mínimo de tres variables, edad, esclerosis de los bordes y matriz son necesarias para la RN de diagnósticos.

5.4. Comparación

La comparación de las áreas bajo las curvas ROC para los modelos de RL y RN en malignidad no fue estadísticamente significativa para malignidad $90,9392$ vs $0,9617$, $p > 0,2$) (Tabla 8, Ilustración 2). Sin embargo, la RN para diagnósticos estableció áreas bajo la curva significativamente mayores que la RL, como demuestran las curvas ROC (Tabla 8, ilustración 2). La sensibilidad y la especificidad de cada modelo en el punto de máxima eficacia diagnóstica de la curva ROC también fueron calculados. No obstante aunque la curva puede dar esos resultados en cada punto de corte, se escogió el punto de máxima eficacia (probabilidad a priori=0,5).

6. Discusión

La importancia de una correcta clasificación en las lesiones de calota es crucial respecto a la actitud diagnóstica y terapéutica. El proceder terapéutico generalmente está determinado por la conducta biológica de las lesiones, y no por otras razones como la sintomatología, ya que la mayoría de las lesiones que se presentan como tumoración, independientemente de su etiología, son resecaadas por motivos estéticos y a petición del paciente (46). En las lesiones malignas se debe obtener un diagnóstico histológico, una intervención quirúrgica o un tratamiento radioterápico, siendo las

lesiones benignas controladas bien mediante estudios de imagen periódicos o clínicamente (47,48). Al ser encontradas habitualmente estas lesiones en los exámenes radio-gráficos, es necesario un correcto enfoque del problema para evitar pérdidas de tiempo, económicas y preocupaciones innecesarias del paciente y personal médico. Este diagnóstico más preciso puede venir por el diagnóstico asistido por ordenador y en concreto, por las redes neuronales.

Las ayudas al diagnóstico que se sirven de instrumentos predictivos basados en reglas y sistemas expertos, hasta el momento no han demostrado un buen rendimiento(49). Sólo recientemente, los modelos basados en redes neuronales han demostrado su superioridad tanto frente al propio diagnóstico humano como respecto a otros diagnósticos por sistemas expertos, añadiendo seguridad en la decisión médica (23,50).

Cuando las radiografías simples eran más frecuentemente realizadas, la incidencia de alteraciones no anatómicas que se presentaban como lesiones líticas de calota era de alrededor de un 7% (1). Sin embargo la incidencia precisa de lesiones de la calota usando TC o resonancia magnética no ha sido estimada. Las distribuciones de las lesiones más comunes en nuestro estudio es algo distinta de la descrita previamente (51), debido a diferentes criterios de selección. Un posible sesgo es que la efectividad del presente diseño de RN debería ser probado en un conjunto más general de lesiones (1,2). Las lesiones óseas fueron predominantemente solitarias, y distribuidas principalmente en el hueso parietal, seguido del frontal. Estos hallazgos son similares a los descritos en cualquier lesión de calota, debido a la mayor superficie del hueso parietal, comparado con los demás huesos del cráneo (1).

Hay acuerdo general de que las apariencias de la mayoría de lesiones de calota en TC y RM son difícil de interpretar. Aunque algunos autores informan que la mayor parte de lesiones son predecibles según su distribución de edad, el diagnóstico radiológico

de éstas es generalmente incierto debido a apariencias radiológicas y clínicas similares (52). Al igual que otras lesiones en el esqueleto (53), existe una clara necesidad para un diagnóstico exacto. Las implicaciones de una mala clasificación hacen útil una ayuda al diagnóstico para tipificar estas lesiones.

Las reglas de predicción clínicas son útiles en varios campos, incluyendo el radiodiagnóstico, como ayuda en el proceso de decisión médica (54). entre los varios métodos para el diagnóstico asistido por ordenador de lesiones óseas, uno de los pioneros fue G Lodwick con su aproximación bayesiano (19). Los principales inconvenientes de este método es no aprende de casos nuevos, requiere un gran número de patrones para proporcionar una matriz probabilística estadísticamente válida (955) y que funciona mejor con enfermedades comunes (56). Entre las distintas técnicas, la RL ha resultado como el análisis estadístico de elección para predecir eventos dicotómicos (27). Recientemente, el uso de análisis vectorial (958) y redes neuronales (3) ha sido utilizado para el reconocimiento de patrones radiológicos (58,59). Uno de los inconvenientes argüidos contra el uso generalizado de RN es que sus resultados son raramente comparables con métodos estadísticos para establecer su rendimiento relativo (57). El mejor método para determinar si las RN son mejores para un problema determinado es comparar con un procedimiento estadístico en los mismos grupos de entrenamiento y validación (57).

Nuestros resultados muestran que una simple RN de tres capas con propagación retrógrada puede ser entrenada para distinguir prácticamente los patrones de lesiones de calota asociadas con benignidad, malignidad y las lesiones más comunes del calvario. Los similares resultados de la RN y la RL para diferenciar lesiones benignas de malignas se debe probablemente a la buena adaptación del modelo estadístico al tamaño de la muestra. Las variables más relevantes seleccionadas tanto por la RL como por la RN fue la edad y la definición

de los bordes. La edad es la variable clínica más importante en el diagnóstico de las lesiones óseas focales. Edeiken (6) observó una mayor frecuencia de lesiones malignas a partir de los 40 años afirmando que con la edad se predice correctamente el 80% de las lesiones malignas óseas (55). La definición de los bordes es una de las variables radiológicas clásicas, desde las descripciones de Lodwick (60), que determinan la agresividad de las lesiones, que mejor definen la conducta biológica de una lesión ósea y resulta extremadamente útil en la aproximación diagnóstica a estas lesiones. Sin embargo, en un estudio de lesiones líticas de calota, la definición de los bordes no fue importante para el diagnóstico, aunque ningún método estadístico fue empleado (1).

En el grupo de diagnósticos histológicos, según disminuía la muestra de pacientes, descendía la capacidad predictiva de la ecuación de RL. Dado que todos los métodos de clasificación dependen del entrenamiento, mejores resultados deben ser esperados según la muestra sea mayor. Por esta razón, los diagnósticos con menos de diez casos fueron descartados (3). Sorprendentemente, la RN se comportó extremadamente bien comparada con la RL incluso en los grupos más pequeños. El método leave-one-out, usado para generalizar el análisis de RL, puede ser el responsable de la baja rentabilidad del modelo estadístico.

Se necesitaron cuatro variables para el diagnóstico conjunto de malignidad y de la histología: edad, definición de los bordes, esclerosis marginal y matriz. Estas variables son conocidas como de gran relevancia en la clasificación de lesiones óseas (55). Está más allá del objetivo de este estudio describir cuales son las variables individuales más importantes para cada diagnóstico. De hecho, no se construyeron diferentes RN para cada enfermedad, aunque éste, sí que fue uno de los inconvenientes de la RL, la cual requirió diferentes ecuaciones y variables para cada una de las histologías.

La superioridad de la RN comparada con la RL refleja la habilidad de la red neuronal para descubrir relaciones complejas entre las variables. Esta tarea es conseguida por la RN debido a su arquitectura paralela, una diferencia cardinal con los métodos estadísticos (17). La forma de procesamiento no lineal de la red neuronal permite incluso, con datos solapados o escasos para su análisis estadístico, encontrar formas de independizarlos y asociarlos a las lesiones que típicamente lo presentan.

El mejor rendimiento de la RN no creemos que se deba a sobreentrenamiento, dado que los métodos utilizados en su diseño fueron específicamente escogidos para evitar este sesgo. Es más, aunque el sobreentrenamiento es una de las explicaciones comúnmente asumidas para el mejor rendimiento de las RN, los métodos estadísticos son también propensos al sobre ajuste (57).

Un artículo previamente publicado describía el análisis con RN de lesiones óseas focales de forma general en todo el esqueleto pero carecía de comparación con un método estadístico (6). Ese trabajo fue ligeramente más sensible que el nuestro, 89,8% vs 86,6% pero marcadamente menos específico 75% vs 90,3%. La razón de estas diferencias puede radicar en que su serie estaba basada en toda clase de lesiones esqueléticas exclusivamente estudiadas con radiografías simples. Dado que las radiografías simples no pueden demostrar todas las características de una lesión, varios de los hallazgos de nuestro estudio obtenidos con TC, no fueron considerados en su trabajo. Los mencionados autores utilizaron validación cruzada para el entrenamiento de su RN, con un 75% de los pacientes utilizados para el grupo de entrenamiento y el restante 25% para prueba; esta aproximación fue intercambiada hasta que toda la muestra fue utilizada para probarla. Esto implicaba que para enfermedades con poca casuística, el número de elementos necesarios para entrenar la red era tan pequeño en el conjunto de entrenamiento que los

pacientes en el grupo de prueba obtenían un muy bajo rendimiento.

El principal problema en clasificar lesiones óseas, incluidas las del cráneo, es la considerable superposición de hallazgos de imagen (61) haciendo inútil la mayoría de las técnicas empleadas para su clasificación, como el análisis vectorial (58). Las RN poseen ventajas potenciales sobre los métodos tradicionales multivariantes para la clasificación de datos sin límites definidos y no lineales entre diferentes clases. La superioridad de las RN sobre las aproximaciones estadísticas parece clara en el reconocimiento de patrones complejos tales como los representados en el presente estudio, incluso con pequeños conjuntos de datos.

7. Conclusiones

Tras la presentación de este trabajo de investigación, se resumen los distintos hallazgos y juicios con las siguientes conclusiones:

1. Las lesiones de la calota suponen un total del 4,16% de las tomografías computarizadas realizadas en un hospital universitario de referencia.
2. Las lesiones benignas suponen un total del 73,05% de las lesiones de la calota, siendo el resto, un 26,95%, malignas. Las lesiones más comunes son las histiocitosis de células de Langerhans (18,6%), osteoma (13,2%), quistes epidermoides (13,2%), metástasis (12,6%), meningioma (10,2%), hemangioma (9%) y displasia fibrosa (6%).

3. El principal problema de la clasificación de las lesiones óseas de la calota es el solapamiento de sus manifestaciones clínicas y radiográficas.

4. La regresión logística logra unos resultados correctos cuando existe un número adecuado de ejemplos y una separación lineal, como es en las lesiones benignas y malignas. La red neuronal fue mejor que la regresión logística, pero sin diferencias estadísticamente significativas.

5. Los rendimientos de la regresión logística para las distintas histologías son menores del obtenido para benigno y maligno. Para el diagnóstico de las lesiones, las redes neuronales se muestran significativamente superiores a la regresión logística.

6. Las redes neuronales descubren interacciones entre las variables independientes no discernibles por medio del análisis estadístico, permitiendo una mejor adaptación a los distintos diagnósticos y clasificando correctamente a la mayoría de estos pacientes.

7. Los patrones radiológicos clásicamente asociados con determinados diagnósticos histológicos, fueron seleccionados por la red neuronal como las variables más importantes. Sin embargo, estas variables no pudieron ser identificadas generalmente en el análisis estadístico.

8. Las redes neuronales pueden utilizarse eficazmente como ayuda al diagnóstico radiológico asistido por ordenador para las lesiones de la calota observadas mediante tomografía computarizada.

Bibliografía

1. Thomas J, Baker H. Assessment of roentgenographic lucencies of the skull: a systematic approach. *Neurology* 1975;25:99-106.
2. Steinmeier R, Huk W, Fahlbusch R. Tumoren und raumfordernde Fehlbildungen der Schädelkalotte. *Radiologie* 1988;28:134-40.
3. Reinus W, Wilson A. Quantitative analysis of solitary lesions of bone. *Invest Radiol* 1995;30:427-32.

4. Selker H, Griffith J, Patil S, et al. A comparison of performance of mathematical predictive methods for medical diagnosis: Identifying acute cardiac ischemia among emergency department patients. *J Invest Med*. 1995;43:468-76.
5. Keats T. Four normal anatomic variations of importance to radiologists. *AJR Am J roentgenol* 1957;78:89-94.
6. Edeiken J. Tumores óseos y estados de tipo

- tumoral. En: Edeiken J, editor. Diagnóstico radiológico de las enfermedades de los huesos. 2ª ed. Buenos Aires: Panamericana; 1991; 4, p. 30-419.
7. Greenspan A. Radiologic evaluation of tumors and tumor-like lesions. En: Greenspan E, editor. Orthopedic Radiology. A practical approach. 2nd ed. New York: Raven Press; 1992; 15, p. 15.1-15.39.
 8. Rafal R, Markisz J. Aplicaciones musculoesqueléticas de la TC. En: Markisz J, editor. Diagnóstico por imágenes en el sistema musculoesquelético. Barcelona. Edika-Med; 1993; 7,p.211-29.
 9. Levine E, Lee K, Neff J, et al. Comparison of computed tomography and other imaging modalities in the evaluation of musculoskeletal tumors. Radiology 1979;131:431-7.
 10. Sundaram M, McGuire M. Computed tomography or magnetic resonance for evaluating the solitary tumor or tumor-like lesion of bone? Skeletal Radiol 1988; 17:393-401.
 11. Kido D, Gould R, taati F, et al. Comparative sensitivity of CT scans, radiographs and radionuclide bone scans in detecting metastatic calvarial lesions. Radiology 1978;128:371-5.
 12. Fryback D, Thornbury J. The efficacy of diagnostic imaging. Med Decis Making 1991;11:88-94.
 13. Kassirer J. our stubborn quest for diagnostic certainty. A cause of excessive testing. N Engl J Med. 1989; 320:1489-91.
 14. Clayton, P. editor. Field trials of medical decision-aids: potential problems and solutions. 1st.Bethesda, USA: American Medical Informatics Association; American Medical Informatics Association; 1991;3p. Proceedings of Fifteenth Annual Conference on Computer Application in Medical Care.
 15. Shortliffe E. Computer programs to support clinical decision making. JAMA 1987;258:61-6. Barnett G, Cimino J, Hupp J, et al. DX plain. An evolving decision-support system. JAMA 1987; 258:67-74.
 17. Taylor P. Inveted reiew: computer aids for desicion-making in diagnostic radiology-a literature review. Br J. Radiol 1995;68:954-57.
 18. Tourassi G, Floyd C, Sostman H, et al. Acute pulmonary embolism: artificial neural network approach for diagnosis. Radiology 1993; 189:555-8.
 19. Lodwick G. Computers in radiologic diagnosis. Appl Radiol 1986; (En/Feb):61-5.
 20. Baxt W. Complexity, chaos and human physiology: the justification for non-linear neural computational analysis. Cancer Lett 1994;77:85-93.
 21. Ryback R, Eckardt M, Rawlings R, et al. Quadratic discriminant analysis as an aid to interpretative reporting of clinical laboratory tests. JAMA 1982; 248:2342-5.
 22. Shufflebarger C. What is a neural network? Ann Emerg Med 1992;21:63-4.
 23. Baxt W, skora j. Prospective validation of artificial neural network trained to identify acute myocardial infarction. Lancet 1996; 347:12-5.
 24. Astion M, Bloch D, Wener M. Neural networks as expert systems in rheumatic disease diagnosis: artificial intelligence opr intelligent artificie? J Rheum 1993; 20:1465-8.
 25. Taube A, tholander B. Over- and underestimation of the sensitivity of a diagnostic malignancy test due to various selections of the study population. Acta Oncol 1990;29:1-5.
 26. Osborn A. Meningiomas and other nonglial neoplasms. En: Osborn A, editor. Diagnostic Neuroradiology. St. Louis: Mosby-Year Book, Inc. 1996;14,p.579-625.
 27. Mahmood A, Caccamo D, Tomecek F, et al. Atypical and malignant meningiomas: a clinipathological reiew. Neurosurgery 1993;33:955-63.
 28. Jenning PG, Cook PL. Bony metastases from intracranial meningioma. Br J Radiol 1983;56:421-2.
 29. Partington MD, Scheithauer BW, Piepgras DG. Carcinoembryonic antigen production associated with an osteolytic meningioma. Case report. J Neurosurg 1995; 82:489-92.
 30. Kinjo T, Al-Mefty O, Kanaan I. Grade zero removal of supratentorial convexity meningiomas Neurosurgery 1993;33:394-9.
 31. Al- Mefty, O. 1996;Comunicación personal.
 32. Golfieri R, Baddeley H, Pringle J, et al. Primary bone tumors: MR morphologic appearances correlated with pathologic examinations. Acta Radiol 1991;32:390-8.
 33. Tubbs W, Brown L, Beabout J, et al. Benign giant cell tumor of bone with pulmonary metastases: clinical findings and radiologic appearances of metastases in 13 cases. AJR Am J Roentgenol 1992;158:331-4.
 34. Greenfield G, Arrington J. Primary malignant tumors. En: Greenfield G, Arrington J, editores. Imaging of bone tumors: A multimodality approach. Philadelphia: J.B.lippincott Co. 1995;2,p43-166.
 35. numerow L, Krcek J, Tranmer B, et al. Growing skull fracture simulating a rounded lytic calvarial lesion. AJNR Am J Neuroradiol 1991;12:783-4.
 36. Press SJ, Wilson S. Choosing between logistic

- regression and discriminant analysis. *J Am Stat Assoc.* 1987;73:699-705.
37. Hinton G. How neural networks learn from experience. *Sci Am* 1992;207:144-51.
38. Kattan M, Becj J. Artificial neural networks for medical classification decisions. *arch Pathol Lab Med* 1995;119:672-7.
39. Miller A, Blott B. Review of neural network application in medical imaging and signal processing. *Med Biol Eng Comput* 1992;30:449-64.
40. Sarle W. ai-fag/neural-nets/part3. Generalization. Available URL: ftp/ftp. sas.com/pub6neural6FAQ3.html 1996;19(December).
41. Ricard M, Lippmann R. Neural network classifiers estimate Bayesian a posteriori probabilities. *Neural Comput* 1992; 4:461-83.
42. Metz CE. Basic principles of ROC analysis. *Semin nucl Medic*, 1978;8:283-98.
43. Bloch D, Moses L, Michel B. Statistical approaches to classification. methods for developing classification and other criteria rules. *Arthritis Rheum* 1990;33:1137-44.
44. Hanley J, mcneil B. The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve. *Radiology* 1982;143:29-36.
45. Metz C, Wang p, Kronman H, Deconinck F, editor. Information processing in medical imaging. The Hague: Nijhoff; 1984; A new approach for testing the significance of differences between ROC curves from correlated data. p.432-45.
46. Arana E, Latorre FF, Revert A, et al Intradiploic epidermoid cysts. *Neuroradiology* 1996;38:306-11.
47. Taveras JM, Wood EH. The Skull. En: Taveras JM, Wood EH, editores. *Diagnostic Neuroradiology*. 2nd ed. Baltimore: Williams & Wilkins; 1977;1,.p.1-230.
48. Schajowicz F. Lesiones seudotumorales. En :Schajowicz F, editor. *tumores y lesiones seudotumorales de huesos y articulaciones*. 1ª ed. Buenos Aires: Panamericana; 1982;9,p.422-520.
49. Wasson J, Sox H, Neff R, et al. Clinical prediction rules. applications and methodological standards. *N Engl J Med* 1985;313:793-9.
50. Baker JA, Kornguth PJ, Lo JY, et al. Breast cancer: prediction with artificial neural network based on BI-RADS standardized lexicon. *Radiology* 1995;196:817-22.
51. Ruge JR, Tomita T, naidich TP, et al. Scalp and calvarial masses of infants and children. *Neurosurgery* 1988;22:1037-42.
52. Young H, Rothman M, Zoarski G, et al. Imaging of transcalvarial masses. *Radiology* 1994;193 Suppl.1 (Abstract):455.
53. Di Rocco C, Jannelli A, Fileni A, Moschini M. Solitary osteolytic lesions of the skull vault in children. *Neurochirurgia Stuttgart* 1982;25:57-61.
54. D'Orsi C, Getty D, Swets R. Reading and decision aids for improved accuracy and standardization of mammographic diagnosis. *Radiology* 1992;184:619-622.
55. Lodwick G. *Radiologic Concepts*. En: Lodwick, G y Hodes P, editores. *the Bones & Joints. An Atlas of Tumor Radiology*. Chicago: Year Book Medical Publishers. Inc. 1973; 1,p.1-82.
56. de Dombal F. Computer-aided diagnosis and medical decision support are not synonymous *Meth Inform Med* 1995;34:361-8.
57. Tu J. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *j Clin Epidemiol* 1996;49:1225-31.
58. Reinus W, Wilsopn A. Quantitative analysis of solitary lesions of bone. *Invest Radiol* 1995;30:427-432.
59. Baker JA, Kornguth PJ, Lo JY, Floyd CE, Jr. ASrtificial ANN: improving the quality of breast biopsy recommendations. *Radiology* 1996;198:131-135.
60. Lodwick G, Wilson A, Farrell C , et al. Determining growth rates of focal lesions of bone from radiographs. *Radiology* 1980; 134:577-83.
61. Greenfield G, Arrington J. Methodology and pathology of bone tumors. En : Greenfield G, Arrington J, editores. *Imaging of bone tumors: A multimodality approach*. Philadelphia: JB Lippincott Co. 1995; 1,p. 1-41.

Variables	Valores
Edad	Años
Sexo	Masculino, femenino
Primer síntoma registrado	Tumor,dolor,cefalea,hallazgo incidental,otros.
Duración de los síntomas	Meses
Número de lesiones	Número
Hueso	Frontal,parieta, occipital,suturas o fontanelas
Centro	Tabla externa, diploe,tabla interna
Diámetro	En milímetros
Forma	Circular, ovoide (imagen en el plano de mayor diámetro)
Carácter	Lítico permeativa, lítico moteada, lítico geográfico, blástica, mixta blástico moteado, mixta blástica geográfica.
Expansividad	No, leve, moderada,severa.
Masa de partes blandas	Ninguna, intracraneal, subgaleal, ambas
Definición de los bordes	Pobre, moderada, bien
Lobulación	Lobulados, lisos
Esclerosis marginal	No, parcial,anillo(<2mm),banda(>2mm)
Reacción perióstica	No, sí (cualquier forma)
Matriz	No, en vidrio esmerilado, calcificada, osificada y secuestro
Afectación cortical	Diploe, interna, externa, ambas corticales
Forma de afectación cortical	No,engrosada, adelgazada,rota.

Tabla 1. Descripción de las variables utilizadas

Lesiones	Núm.	%	Lesiones	Núm.	%
Histiocitosis de células de Langerhans	31	18.6	Fibroma	2	1.2
Osteoma	22	13.2	Osteomielitis	2	1.2
Quiste epidermoide	22	13.2	Quiste óseo aneurismático	2	1.2
Metástasis	21	12.6	Fascitis nodular	2	1.2
Meningioma	17	10.2	Otros malignos	7	4.2
Hemangioma	15	9	Carcinoma epidermoide cutáneo	3	1.8
Displasia Fibrosa	10	6	Osteosarcoma	2	1.2
Otros benignos	22	13.2	Plasmocitoma	1	0.6
Quiste dermoide	4	2.4	Sarcoma dural	1	0.6
Angioma cutáneo	4	2.4	Total lesiones benignas	122	73.05
Exostosis	3	1.8	Total lesiones malignas	45	26.95
Neurofibroma plexiforme	3	1.8			

Tabla 2. Diagnósticos histológicos, n=167

Diagnósticos	n	%	Rango de edad(años)	Mediana
Histiocitosis de cél.L	31	18.6	0.5-35	4
Osteoma	22	13.2	5-84	51.5
Quiste epidermoide	22	13.2	1-80	25.5
Metástasis	21	12.6	5-80	62
Meningioma	17	10.2	29-74	51
Hemangioma	15	9	20-63	44
Displasia fibrosa	10	6	8-64	18
Otros benignos	22	13.2	0.5-34	14
Otros maligno	7	4.2	16-80	43

Tabla 3. diagnósticos histológicos más frecuentes con su distribución por edad.

Variabes	OR	CI95%	p
Edad	1,0748	1,04-1,10	<0,001
Borde (vs. pobre)			
Moderado	0,2	0,03-1,15	0,07
Bien definido	0,011	0,001-0,09	<0,001
Masas partes blandas (vs. no)			
Intracraneal	43,03	4,06-455,1	0,002
Subgaleal	10,94	0,99-120,7	0,05
Ambas	21,63	1,75-267,1	0,01
Número de lesiones	6,2014	0,97-39,48	0,05

Tabla 4. Variables incluidas en la ecuación de regresión logística para el diagnóstico de malignidad. OR:odds ratio (razón de posibilidades), CI:intervalo de confianza.

Diagnósticos	Variables	OR	IC95%	p
Hiscitiositosis de Células de Langerhans	Lobulación (vs. lobulado)			
	Liso	0,072	0,15-0,33	<0,001
	Forma (vs.circular)			
	Ovoide	0,226	0,07-0,71	0,011
	Edad	0,886	0,84-0,93	<0,001
Osteoma	Centro (vs.diploe)			
	Tabla externa	174,5	12,23-2490	<0,001
	Tabla interna	34,85	0,96-1263	0,053
	Edad	1,05	1,01-1,09	0,003
	Diámetro	0,93	0,88-0,98	0,017
	Duración de los síntomas	1,02	0,99-1,04	0,058
Quiste epidermoide	Hueso (vs.Frontal)			
	Parietal	0,42	0,13-1,35	0,149
	Occipital	1,8	0,56-5,70	0,318
	Suturas/fontanelas	12	0,98-146,5	0,052
Metástasis	Borde (vs. pobre)			
	Moderado	0,09	0,02-0,38	<0,001
	Bien definido	0,01	0,01-0,11	<0,001
	Edad	1,06	1,02-1,10	<0,001
	Numero de lesiones	4,36	1,10-17,28	0,036
Meningioma	Sexo (vs.Masculino)			
	Femenino	7,40	1,64-33,28	0,009
	Reacción perióstica (vs.no)			
	Sí	20,72	5,07-84,7	<0,001
Hemangioma	Forma (vs. circular)			
	Ovoide	0,17	0,04-0,65	0,01
	Masa de parte blandas (vs.no)			
	Intracraneal	0,03	0,003-0,36	0,005
	Subgaleal	0,21	0,04-1	0,051
	Ambas	1,10	0,18-6,76	0,912
	Reacción perióstica (vs.no)			
	si	9,79	1,65-57,96	0,012
Displasia fibrosa	Diámetro	1,03	1,01-1,06	0,003
	Edad	0,96	0,93-0,99	0,038

Tabla 5. Variables incluidas en las distintas ecuaciones de regresión logística para diagnósticos. OR: odds ratio, IC: intervalo de confianza.

Orden	Malignidad		Diagnósticos	
	Variables	A(z)	Variables	A(z)
1	Edad	0.9125	Matriz	0.9411
2	Borde	0.9198	Esclerosis marginal	0.9432
3	Matriz	0.9250	Edad	0.9435
4	Masa de partes blandas	0.9262	Masa de partes blandas	0.9484
5	Sexo	0.927	Borde	0.9516
6	Síntomas	0.9287	Reacción perióstica	0.9523
7	Hueso	0.9295	Lobulación	0.9523
8	Expansividad	0.9315	Tipo de afectación cortical	0.9549
9	Forma	0.9318	Forma	0.9550
10	Tipo de afectación cortical	0.9323	Diámetro	0.9551
11	Lobulación	0.9327	Duración de los síntomas	0.9556
12	Centro	0.9333	Número de lesiones	0.9559
13	Carácter	0.9338	Síntomas	0.9561
14	Reacción perióstica	0.9346	Sexo	0.9568
15	Número de lesiones	0.9353	Afectación cortical	0.9581
16	Duración de los síntomas	0.9354	Expansividad	0.9598
17	Afectación cortical	0.9358	Carácter	0.9615
18	Esclerosis marginal	0.9369	Hueso	0.9623
19	Diámetro	0.9379	Centro	0.9629
	ninguna excluida	0.9617	Ninguna excluida	0.9673

Tabla 6. Orden de las variables de entrada mediante exclusión única de las mismas para las redes de maligno y para los diagnósticos. A(z): área bajo la curva ROC.

Diagnósticos	Todas las variables	Matriz	Matriz, esclerosis de bordes	Matriz, esclerosis de bordes y edad
Histiocitosis c. Lang.	0.9854±0.018	0.6385±0.0591, <0.001	0.7949±0.051, <0.001	0.9852±0.0153, NS
Osteoma	0.9998±0.002	0.8808±0.0484, <0.05	0.9431±0.0348, NS	0.9361±0.0368, NS
Quiste epidermoide	0.9961±0.002	0.6347±0.0682, <0.001	0.9396±0.0358, NS	0.9580±0.0302, NS
Metástasis	0.9944±0.003	0.5225±0.0692, <0.001	0.7663±0.0633, <0.05	0.9446±0.0351, NS
Meningioma	0.9771±0.025	0.5921±0.0768, <0.001	0.7297±0.0726, <0.001	0.8580±0.0581, NS
Hemangioma	0.9409±0.004	0.6558±0.0804, <0.001	0.8460±0.0644, NS	0.9409±0.0428, NS
Displasia fibrosa	0.9601±0.041	0.8804±0.071, NS	0.9070±0.0639, NS	0.7937±0.086, NS

Tabla 7. Resultado de la eliminación de las variables según el orden de la tabla anterior. La eliminación de las variables no disminuyó la A(z) hasta que la edad fue eliminada. Entonces la RN mostró una disminución de la A(z) para histiocitosis de células de Langerhans ($p < 0.001$), meningioma ($p < 0.001$) y osteoma ($p < 0.05$) con un descenso no significativo de la A(z) en los restantes diagnósticos. Esto significa que un mínimo de tres variables, edad, esclerosis de los bordes y matriz se necesitan para la RN de diagnósticos.

Grupos	Regresión logística		Red neuronal	
	A(z)±SD	S/E*	A(z)±SD	S/E*
Maligno	0.939±0.051	79.7/95.2	0.951±0.042	86.6/90.3
Histiocitosis c. Lang	0.911±0.071	92.4/46.9	0.985±0.018	95.9/97.8
Osteoma	0.902±0.037	81.5/83	0.999±0.002	98.1/98.1
Quiste epidermoide	0.544±0.078	77.7/12.3	0.996±0.002	99.9/97.8
Metástasis	0.919±0.091	82.7/85	0.994±0.003	98.8/98.8
Meningioma	0.690±0.078	39.9/87.2	0.977±0.025	95.6/98.8
Hemangioma	0.721±0.075	75.4/53.4	0.940±0.004	93/93
Displasia fibrosa	0.837±0.049	98/7.5	0.960±0.041	98/91

Tabla 8. Comparación de regresión logística y red neuronal mediante áreas bajo la curva ROC. A(z): área bajo la curva ROC. SD: desviación estándar. S/E*: Sensibilidad/Especificidad, números expresan porcentajes.

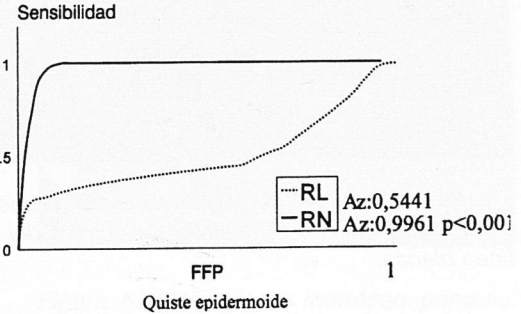
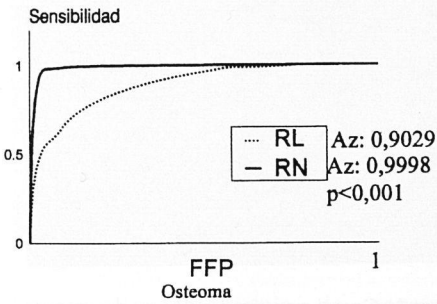
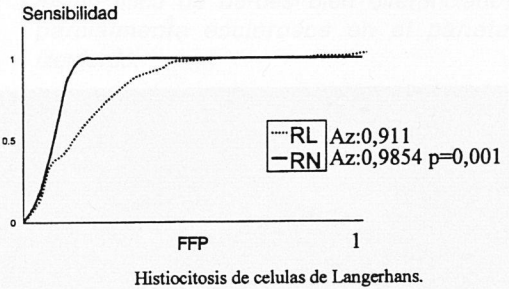
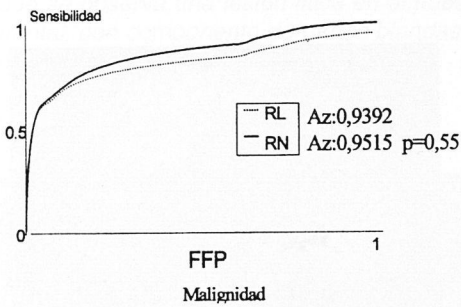


Ilustración 2. Curvas ROC de los distintos diagnósticos FFP: fracción falsos positivos

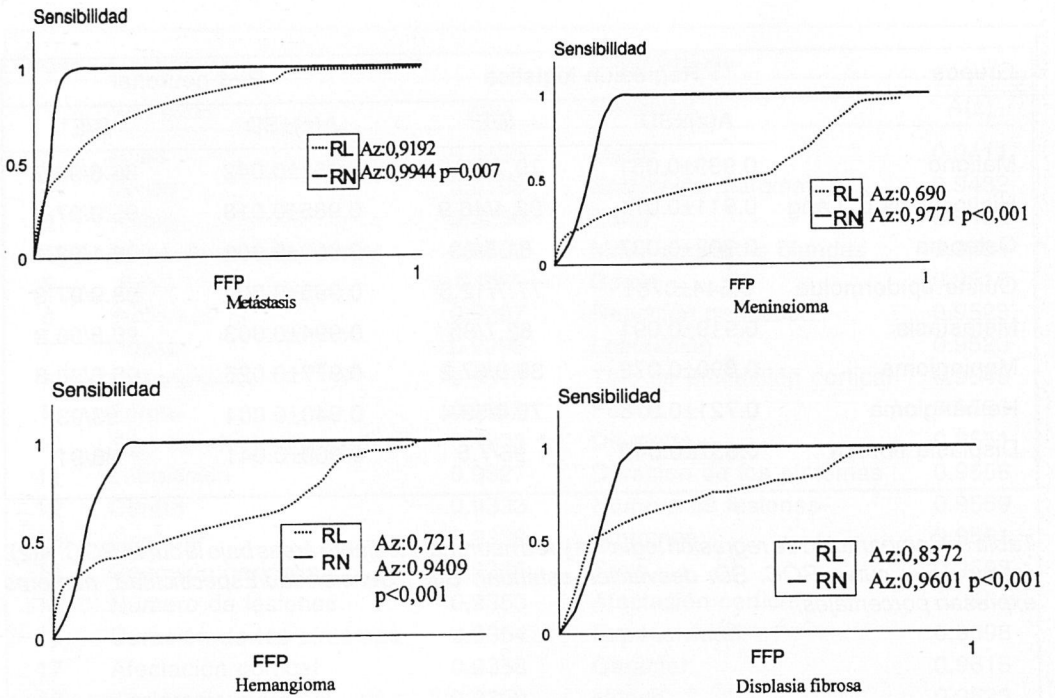


Ilustración 2. Curvas ROC de los distintos diagnósticos FFP: fracción falsos positivos

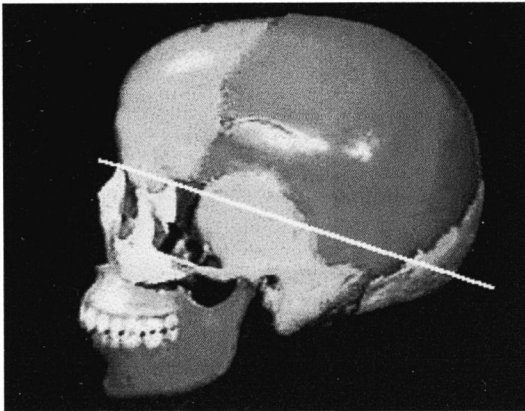


Figura 1. Representación sobre un cráneo coloreado del área de la calota por encima de la línea blanca.

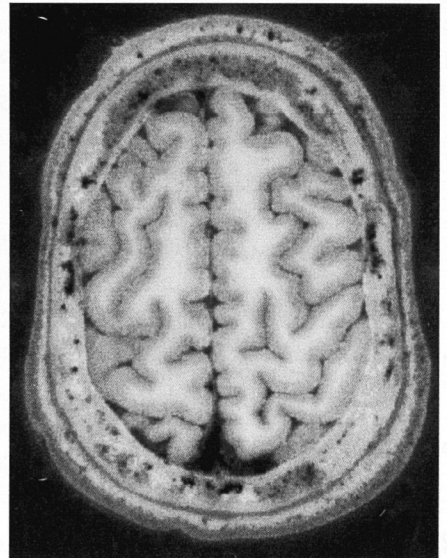


Figura 2. Corte anatómico de cadáver en el que se observan los huesos de la calota con médula ósea roja y grasa, además de espacios diploicos

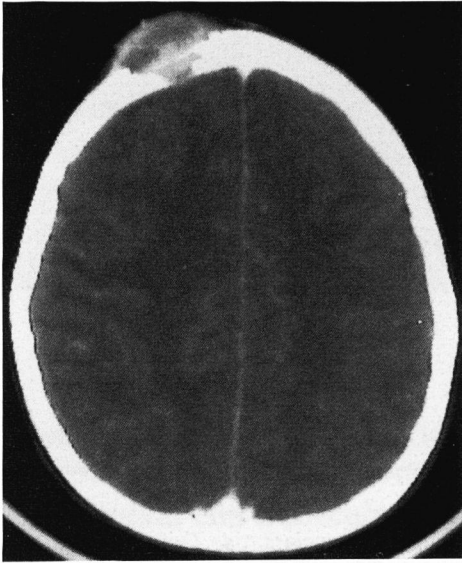


Figura 3. Histiocitosis de células de Langerhans. Imagen axial de TC con contraste, en la que se observa una lesión lítica en el hueso frontal, con componente de partes blandas y con realce heterogéneo en su interior.

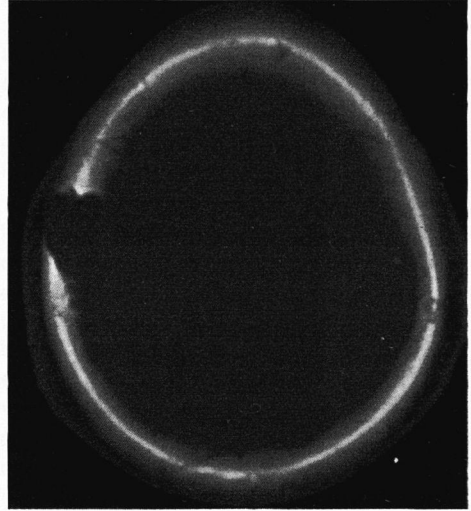


Figura 4. Quiste epidermoide. Imagen axial de TC con ventana ósea en la que se observa una lesión lítica de bordes bien diferenciados parcialmente esclerosos en el parietal izquierdo.

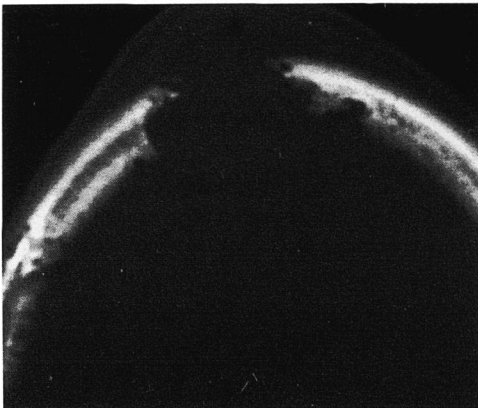


Figura 5. Metástasis de carcinoma hepatocelular. Imagen axial de TC con ventana ósea. Lesión lítica, con bordes mal definidos y componente de partes blandas intra y extracraneal

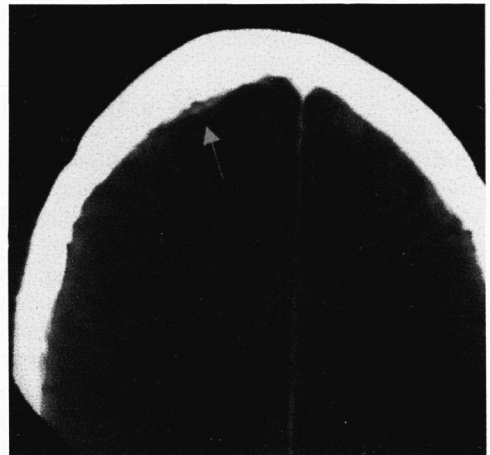


Figura 6. Meningioma intraóseo primario. Imagen axial de TC con contraste. Solo se observa un discreto realce dural sobre la zona de la tomografía referida por el paciente (flecha)

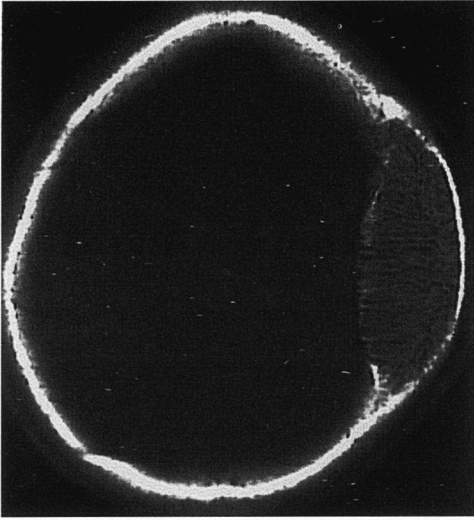


Figura 7. Hemangioma parietal izdo. Imagen axial de TC con ventana ósea. Lesión mixta con trabeculación paralela en su interior localizada en el hueso parietal derecho.

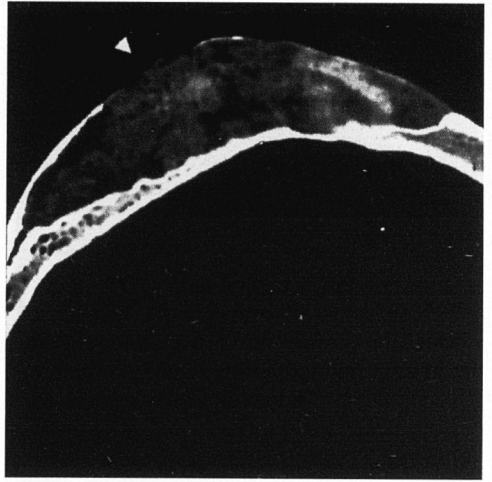


Figura 8. Displasia fibrosa. Detalle de la imagen axial de TC a nivel parietal izquierdo de TC a nivel parietal izquierdo. Nótese el patrón en vidrio esmerilado inhomogéneo y la erosión de la tabla externa (flecha)