

CAPÍTULO 1.

INTRODUCCIÓN.

Desde hace unos años se está asistiendo a la introducción progresiva de los ordenadores en la práctica médica. Aparentemente, la combinación de máquina y programa puede resolver cualquier problema, pero cuando se conecta el ordenador y se pretende hacer una tarea concreta aparecen múltiples dificultades que, casi siempre, dependen de tres factores:

1. El problema más común radica en que no se ha encontrado un sistema informático adecuado a la necesidad que se pretende resolver. Esta situación a menudo se solventa adaptando dicha necesidad al sistema informático disponible. Ello genera insatisfacción y decepciones respecto a las posibilidades de la informática.
2. Otro problema surge cuando existe el sistema informático adecuado pero no está al alcance de nuestras posibilidades económicas, situación no infrecuente en un medio donde, a menudo, se mezclan intereses comerciales excesivos.
3. El último problema se plantea cuando el usuario confunde la capacidad de los ordenadores para procesar datos con la capacidad para organizar y planificar, y se olvida de definir con precisión qué es lo que se pretende resolver mediante el ordenador. Éste es un problema complejo. La organización de la información afecta a todos los aspectos de la sociedad y la medicina no es ajena a esta dinámica. El empleo de los ordenadores y la capacidad de intercambiar información entre ellos mediante redes informáticas permite que las telecomunicaciones sean una realidad que condiciona de manera directa nuestra forma de trabajo.

Las técnicas de procesado de imagen se han desarrollado durante las dos últimas décadas para ayudar al diagnóstico de cáncer de mama. Se ha comprobado que una temprana detección del cáncer de mama puede incrementar la tasa de supervivencia a los cinco años de su aparición de un 60% (aquellos que no han sido detectados precozmente) a un 82%, según el *Instituto nacional del cáncer* (véase referencia 5).

Como consecuencia, durante los últimos años los programas de monitorización de cáncer de mama se han convertido en una herramienta de obligada utilización para los profesionales. Es por ello por lo que los radiólogos deben examinar un importante número de imágenes, muchas de ellas sin ninguna patología, provocando una tasa de errores de entre un 10% y un 30%.

Se hace evidente, pues, la necesidad de la existencia de una herramienta capaz de realizar una segunda lectura de la imagen.

Las herramientas informáticas se han revelado como un poderoso sistema para solucionar el problema antes mencionado, incrementando la sensibilidad del lector en un 10% (véase referencia 3). Uno de los bloques principales de un sistema CAD para cáncer de mama consiste en la **detección de microcalcificaciones**, hecho que resulta crucial para un precoz diagnóstico del cáncer de mama.

Mientras que al principio numerosos estudios se centraban sólo en sistemas restrictivos con la idea de aislar microcalcificaciones del resto de la mamografía, recientemente, se está realizando un importante esfuerzo para incluir redes neuronales, u otros clasificadores, que puedan obtener regiones de decisión sensiblemente más complicadas. Con este objetivo, se debe realizar un paso previo de extracción y selección de características del entorno de la posible microcalcificación. Es en las mamografías densas donde encontramos una dificultad mayor ya que las microcalcificaciones son difíciles de detectar debido al similar valor de intensidad existente en sus alrededores. En estas circunstancias, con el fin de conseguir una alta sensibilidad, se permite un amplio número de falsos positivos.

Con el propósito de decidir si una microcalcificación es maligna o benigna se han propuesto multitud de clasificadores. Sin embargo, esto continúa siendo una tarea ardua incluso para los radiólogos expertos. En general, los resultados obtenidos en los diversos trabajos de investigación realizados en esta área son prometedores ya que han conseguido una alta sensibilidad y especificidad. No obstante, la tasa de falsos positivos necesitan ser disminuidos aún, manteniendo la alta sensibilidad.

Normalmente, los investigadores invierten su esfuerzo en optimizar un bloque en concreto del esquema general de detección, olvidándose por completo de los otros bloques. Se podrían conseguir mejores resultados si se llevara a cabo un estudio general del conjunto de las herramientas CAD. En este sentido, sería del todo deseable el diseño de una herramienta CAD que, escogiendo para cada caso la mejor aplicación diseñada, nos llevara hacia un rendimiento óptimo.

Y con este fin nace este proyecto. Nos proponemos realizar una aplicación que detecte con la mayor fiabilidad posible las microcalcificaciones en las mamografías y que utilice todas las técnicas disponibles en la actualidad. Además nos planteamos el reto de decrementar el coste computacional de herramientas ya diseñadas.

Nuestro estudio está claramente dividido en dos partes. En primer lugar, analizaremos la mamografía y obtendremos unos puntos, *candidatos*, que consideramos como posibles microcalcificaciones. En segundo lugar, haremos un estudio de esos puntos y decidiremos si son realmente o no microcalcificaciones. Pero profundicemos un poco en cada uno de los procedimientos seguidos.

Para la obtención de nuestros puntos candidatos hemos realizado una aplicación bajo el entorno de programación C. La primera de nuestras tareas consistirá en convertir nuestra mamografía a un formato comercial que sea capaz de leer cualquier computador. Así pues, el primer paso que damos consiste en convertir nuestra imagen original de *extensión LUM* a otra de *extensión BMP*. A continuación, haremos un filtrado de la imagen tratando por todos los medios que la parte más oscura de la imagen, es decir, la no perteneciente a la mama, afecte lo menos posible a nuestro procesado. Una vez tenido en cuenta este hecho, aplicaremos un algoritmo de detección de píxeles basado en la predicción lineal de éstos. Finalmente, con el fin de eliminar los numerosos falsos candidatos que se obtienen, se ha definido el *parámetro TR*, cuya aplicación a estos puntos proporcionará un cribado del orden de 60%, lo cual supone un importante ahorro computacional a la hora de utilizar la segunda aplicación implementada.

Como segundo paso, usaremos un descriptor basado en una *máquina de vectores soporte (SVM)*. El descriptor ha sido realizado en el entorno MatLAB y su fundamento teórico radica en los estudios de *Vladimir Vapnik* (véase referencia 15). Para ello, usamos un artificio similar a una red neuronal, esto es, entrenamos al programa con 630 puntos singulares en los que le decimos, dadas las características de los píxeles que rodean a dichos puntos, si debe considerar o no como microcalcificaciones a los candidatos que le pasamos como parámetro.

Sin más preámbulos, pasemos a describir cada uno de los puntos que se han mencionado en esta introducción.