

Análisis de la densidad de mama asistido por ordenador

Trabajo de investigación del Master de Inteligencia Artificial,
Reconocimiento de Formas e Imagen Digital

Departamento de Sistemas Informáticos y Computación,
Universidad Politécnica de Valencia

Dirigido por:
Rafa Llobet y Roberto Paredes

Presentado por:
Juan Antonio Solves Llorens

1. Resumen.....	3
2. Introducción	4
2.1. Radiografías y rayos X.....	4
2.2. Cáncer de mama.....	9
2.3. Estudio de la densidad de mama	11
2.4. Clasificación asistida por ordenador	14
3. Clasificación de la densidad de mama asistida por ordenador	16
3.1. Segmentación de mama	16
3.2. Clasificadores.....	20
3.3. Estado del arte.....	24
3.4. Aproximación propuesta	29
4. Experimentación	31
4.1. Preprocesado de la imagen.....	32
4.2. Extracción de características y experimentación	36
4.3. Resultados con histograma global normalizado.....	37
4.4. Resultados al aplicar diversas técnicas sobre el histograma global	40
4.5. Resultados con técnicas aplicadas sobre ventanas locales.....	48
4.6. Estudio gaussiano de la densidad.....	54
5. Conclusiones.....	60
6. Apéndices.....	62
6.1. Base de datos MIAS.....	62
6.2. Base de datos DDSM	62
6.3. Recursos utilizados	63
7. Bibliografía	64

1. Resumen

En el presente trabajo se realiza un estudio sobre la densidad de la mama y diversos experimentos asistidos por el ordenador para clasificar mamografías según su densidad. Varios estudios han demostrado que existe una relación directa entre la densidad de mama y el riesgo de padecer cáncer de mama, de manera que una herramienta eficiente que clasifique la densidad de las mamografías permitiría a los expertos en detección de cáncer centrarse en las mamografías de más riesgo.

El estudio de la densidad de la mama presenta algunos problemas. En primer lugar, la subjetividad con la que un experto humano clasifica una mamografía puede originar que se disponga de varias clasificaciones para una misma imagen, dependiendo del experto que la haya etiquetado. La falta de una base de datos amplia y consensuada hace que comparar resultados entre diferentes estudios sea complicado ya que muchos de ellos usan conjuntos de mamografías privados y de acceso restringido, y otros usan subconjuntos sin especificar de los repositorios de mamografías públicos. A estos inconvenientes hay que añadirle el problema propio del análisis de imágenes, ya que diferenciar entre varios niveles de gris para determinar la densidad de una imagen no es algo sencillo.

A lo largo del trabajo, se muestra la experimentación seguida en el estudio de la densidad. En primer lugar se intentarán varias aproximaciones utilizando el histograma global de las mamografías. Después se probarán técnicas más complejas sobre el histograma para tratar de mejorar la clasificación y se cerrará la experimentación con un planteamiento de histogramas a partir de ventanas locales de la imagen, comparándolo con las metodologías anteriores. Se demostrará de este modo que la aproximación local obtiene mejores resultados que las otras opciones y que es por donde hay que seguir investigando.

El texto está dividido en varios capítulos. En primer lugar se realizará un repaso general a la forma en que se obtienen las mamografías, la actualidad sobre el cáncer de mama, la relación entre la densidad de mama y el cáncer y, por último, un resumen sobre el uso del ordenador para el análisis de la densidad. En el siguiente capítulo del trabajo se expondrá un resumen sobre el estado del arte relacionado con la clasificación de la densidad, en el que también se realizará un breve repaso de los clasificadores más utilizados para tratar este problema. Por último, se expondrán los experimentos realizados y los resultados, apoyados con gráficas y explicaciones de detalles y problemas que se hayan encontrado a la hora de realizarlos. El trabajo finaliza con la presentación de las conclusiones y las posibilidades de trabajo futuro, un apéndice en el que se reseñan las bases de datos y programas utilizados y el listado de bibliografía consultada.

2. Introducción

2.1. Radiografías y rayos X

En general, las imágenes médicas se obtienen a partir de señales que el cuerpo humano emite tras ser afectado de una forma determinada. Por ejemplo, para obtener una radiografía se envían rayos X sobre el paciente. Una parte de los rayos es absorbida por el paciente y otra le atraviesa, llegando al detector para proporcionar la imagen. En la *Tabla 1* se muestran diversas imágenes que se obtienen del cuerpo humano y la forma en que se activa.

IMAGEN	ACTIVACIÓN
Radiografía (RD)	Rayos X
Fluoroscopia (FD)	Rayos X
Angiografía (AD)	Rayos X
Tomografía Axial Computadorizada (TAC)	Rayos X
Resonancia Magnética Nuclear (RMN)	radiofrecuencia y campos magnéticos
Medicina Nuclear (SPECT) y (PET)	radiación γ
Ecografía	ultrasonidos
Ecodoppler	ultrasonidos

Imagen 1. Imágenes médicas y activación que las origina.

Las imágenes médicas pueden presentarse en formato analógico o digital. El formato digital se está imponiendo gracias a los avances tecnológicos y a las ventajas que aporta (entre ellas: el menor espacio físico para su almacenamiento, la facilidad de transporte y consulta, etc.). Dada la naturaleza del presente trabajo, se va a realizar un breve resumen sobre la generación y uso de los rayos X para la obtención de imágenes médicas.

El físico alemán *Wilhelm Conrad Roentgen* fue el descubridor de los rayos X en 1895. El físico observó que un tipo de energía electromagnética (de naturaleza no nuclear) podía penetrar objetos opacos y proporcionar una imagen de su estructura interna. Debido a esta propiedad y a no saber en qué consistía, llamó a este tipo de radiación rayos X. Los valores de energía de los rayos X les permiten arrancar electrones de los átomos (a este proceso se le denomina ionización). Desde entonces, se ha convertido en una de las herramientas más importantes y empleadas en medicina, tanto en diagnóstico como en terapia.

Los rayos X se generan artificialmente para su uso médico mediante tubos de rayos X. Estos tubos generan haces de rayos X mediante la emisión, aceleración e interacción de electrones con un material específico. El tubo contiene un alto vacío en su interior (10^{-6} mm de Hg, aproximadamente) y está formado por dos electrodos:

-Un cátodo, que emite electrones por efecto termoeléctrico al circular electrones por su interior.

-Un ánodo, que acelera a los electrones atrayéndolos (generalmente se trate de tungsteno o wolframio conectado a un potencial positivo superior a 40 kV).

El vacío sirve para impedir que los electrones choquen con otras partículas que no sean el ánodo, aumentando la eficacia del proceso.

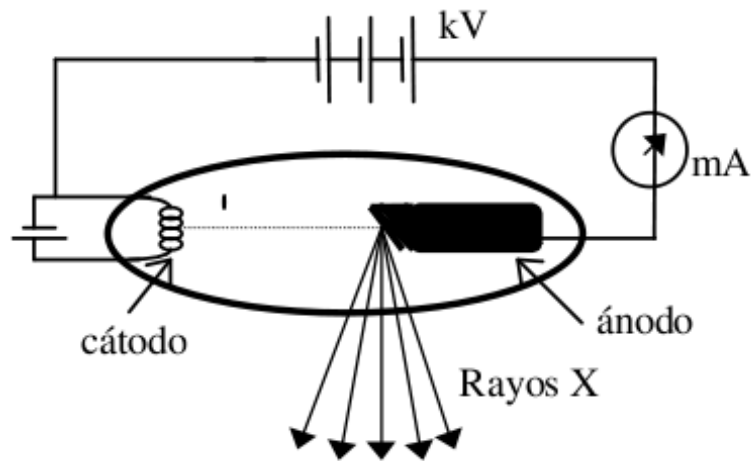


Imagen 2. Tubo de rayos X.

Sus parámetros más relevantes para la aplicación médica son:

- El voltaje, que regula el poder de penetración del haz. Para diagnóstico (radiografías) varía entre 20 y 150 kV, mientras que en terapias (tratamiento de tumores) puede llegar a los 300 kV.
- La intensidad de corriente por el tiempo de emisión. Varía entre 10 y 160 mA por segundo.

Cuando el haz de electrones acelerados interacciona con los átomos del ánodo se genera una radiación electromagnética (fotones) que recibe el nombre de rayos X. Se produce mediante dos mecanismos diferentes:

- Radiación característica: se produce al chocar los electrones acelerados (que vienen del cátodo) con los electrones de las capas más internas de los átomos del ánodo. Éstos saltan de su órbita y su lugar es reemplazado por los electrones de las capas más externas, lo que genera una emisión de radiación característica que son los rayos X cuya energía hf es igual a la diferencia entre los dos niveles energéticos entre los incidentes y la de enlace en la capa correspondiente. La radiación característica de interés en radiología es la que corresponde a la expulsión de los electrones de la capa K, por lo que la energía de los electrones incidentes tiene que ser superior a la de enlace en la capa K (69.4 keV para el Wolframio, por ejemplo).

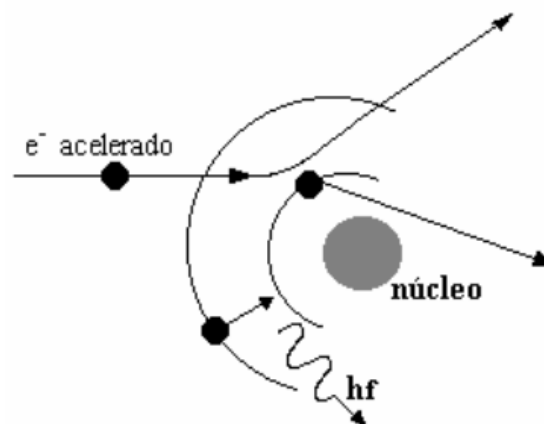


Imagen 3. Radiación característica.

- Radiación de frenado o Bremsstrahlung: cuando el electrón proveniente del cátodo se acerca al núcleo de un átomo del ánodo, es atraído hacia este desviándose de su

trayectoria. El cambio de velocidad (deceleración) producido genera una pérdida de energía del electrón que se emite en forma de radiación de rayos X, llamada de frenado.

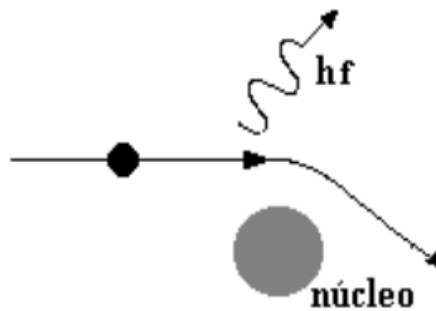


Imagen 4. Radiación de frenado.

Por lo tanto, se genera radiación con energías que corresponden a la radiación característica y la de frenado, constituyendo ambas los rayos X. En la generación de rayos X se producen importantes pérdidas de calor en el tubo, ya que aproximadamente el 1% de los choques produce los rayos X mientras que el 99% restante cede la energía en forma de calor. Por esta razón, es importante que el tubo disponga de un sistema de refrigeración.

En principio, los rayos X pueden salir en cualquier dirección, pero debido a la inclinación del ánodo existe una dirección predominante en la que se encara una ventana para permitir la salida de la radiación. En esta ventana se utiliza un filtro (generalmente de aluminio) para absorber las radiaciones menos energéticas, ya que se necesita un haz homogéneo y de elevada energía para resultar efectivo. La radiación ya filtrada se controla mediante un diafragma que regula su dirección, y desde ahí los rayos X se propagan por el aire hasta el cuerpo sobre el que han de actuar.

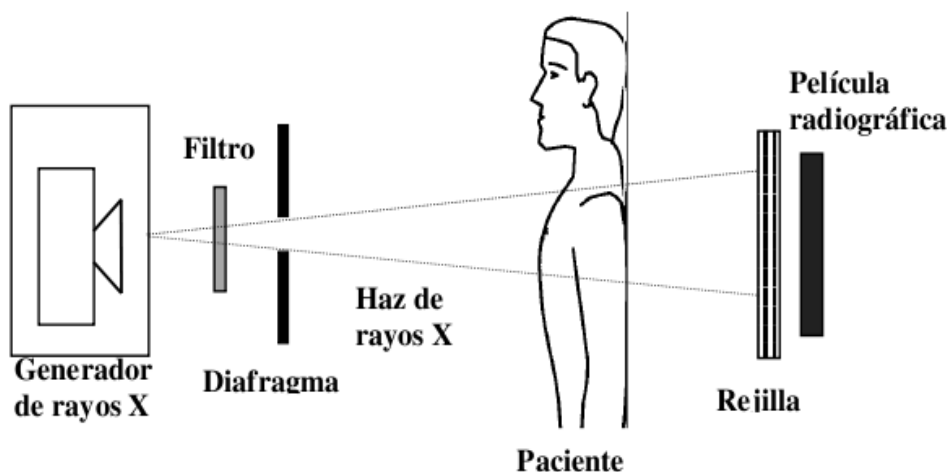


Imagen 5. Sistema de radiografía.

Cuando el haz de rayos X atraviesa un cuerpo, pueden ocurrir dos cosas: que lo atraviesen sin interacción o que interaccionen.

Los rayos X que atraviesan el cuerpo sin interaccionar con él representan aproximadamente el 5 % de la radiación.

El 95 % restante de la radiación interactúa con los electrones de los átomos que forman el cuerpo. La interacción se produce, principalmente, mediante dos mecanismos:

-Efecto fotoeléctrico. Se produce cuando el fotón que incide sobre los átomos del cuerpo transfiere toda su energía a un electrón, arrancándole de su órbita. El fotón incidente desaparece y el vacío dejado en la órbita es ocupado por otro electrón de una capa más externa, lo cual genera otro fotón (pero no es de rayos X).

-Efecto Compton. Se produce cuando el fotón que incide sobre los átomos del cuerpo transfiere parte de su energía (no toda) a un electrón, arrancándole de su órbita. El fotón incidente se desvía de su trayectoria y, como en el efecto fotoeléctrico, se produce la ocupación del “hueco” por otro electrón de una capa superior, emitiéndose otro fotón (que tampoco es de rayos X).

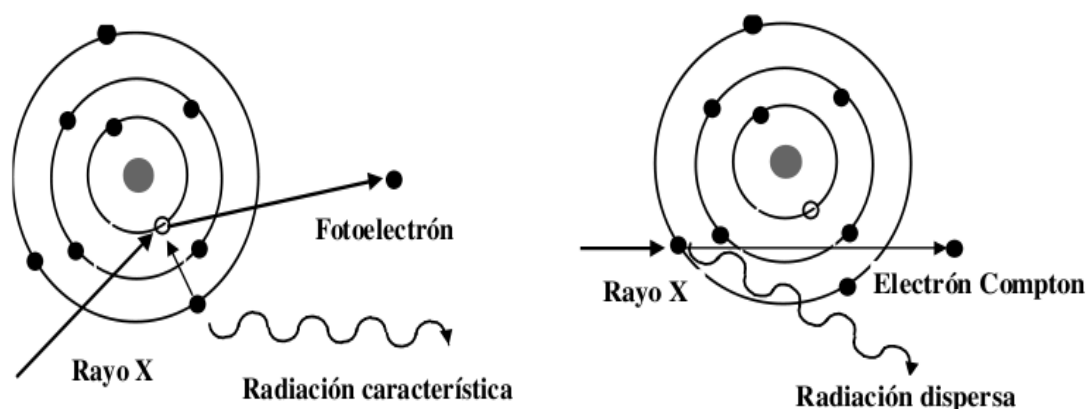


Imagen 6. Efecto fotoeléctrico (izquierda) y Compton (derecha).

La radiación que no interacciona con el cuerpo son los rayos X que se utilizan para impresionar la radiografía. La que interacciona con el cuerpo es la importante. Según el número de interacciones, que depende de la estructura y espesor del material que atraviesa, se produce mayor o menor atenuación del haz. Estas atenuaciones son las que proporcionan los diferentes niveles de gris en las radiografías. Los fotones que no son de rayos X resultantes de las interacciones se eliminan, interponiendo una rejilla difusora antes de que alcancen la placa de impresión (Imagen 5).

La energía del haz debe ser tal que la atenuación de la radiación en la zona del cuerpo sea selectiva para las distintas estructuras que haya en su interior. El haz de salida tendrá distinta intensidad en los distintos puntos y al incidir sobre una pantalla fluorescente o una placa fotográfica permitirá ver la imagen correspondiente. Por ejemplo, cuando los rayos X incidan sobre la pantalla o la placa tras atravesar un hueso la atenuación habrá sido mayor y producirá poca luminosidad, mientras que los haces que hayan atravesado partes blandas habrán sufrido menos atenuación y llegarán con una mayor intensidad, dejando una luminosidad en la placa mucho mayor.

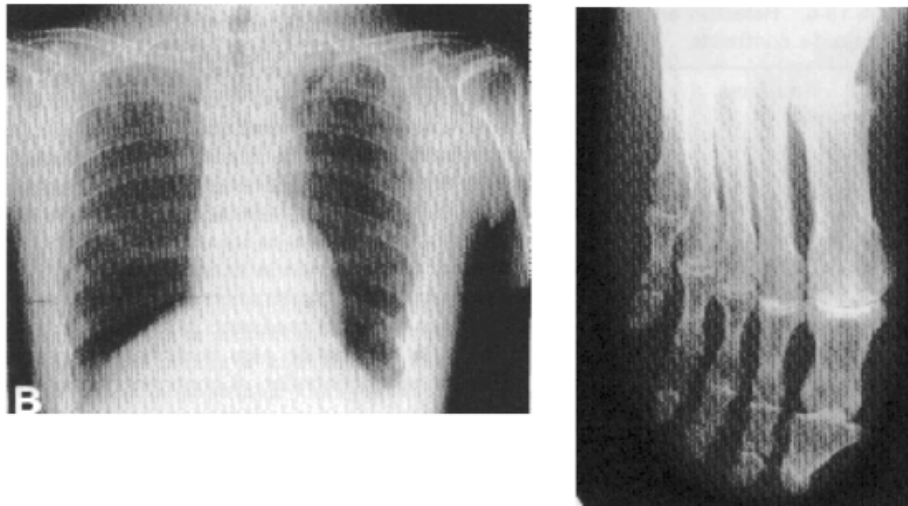


Imagen 7. Radiografías.

La obtención de radiografías en formato digital, para su posterior estudio y tratamiento informático se obtiene de dos formas: digitalizando las radiografías convencionales u obteniéndolas directamente en formato digital. Para esta segunda opción se usan placas fosforescentes fotoestimulables. El receptor de la imagen es una placa con fluorohaluros de Bario sensibles a los rayos X. La imagen se forma mediante trampas de alta energía, que capturan electrones de valencia. Para hacer visible la imagen, se emplea un láser para estimular a los electrones capturados y que vuelvan a su banda de energía, emitiendo luz azul. Esta luz azul es detectada por un tubo fotomultiplicador, que emite señales eléctricas. Las señales se digitalizan y almacenan para ser visualizadas en un monitor o impresas. Estos sistemas de radiografía digital son muy costosos y caros, estando aún en investigación.

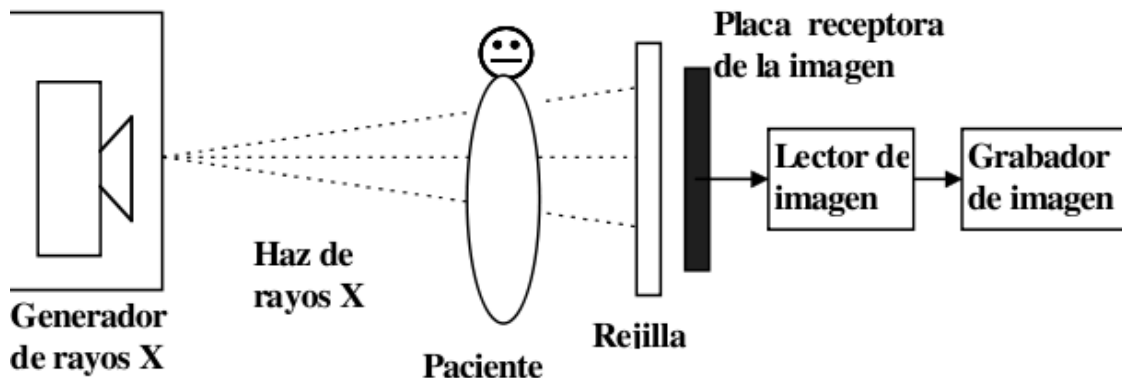


Imagen 8. Sistema de radiografía digital.

A lo largo de los años se ha podido observar que la radiación ionizante daña los tejidos. Sería positivo establecer un umbral máximo de exposición a partir del cual el tejido resulta dañado, pero ese nivel no existe o es extremadamente difícil de determinar. El daño que puede causar depende del tiempo de exposición, es decir, una misma cantidad produce menos efectos dañinos dividida en varias exposiciones que en una única exposición. También el tamaño del área expuesta influye en el tamaño del daño que se pueda producir, ésta es la razón por la que en radioterapia la radiación se dirige a una zona en concreto mientras que se aísla el resto del cuerpo. La sensibilidad varía entre especies y dentro de una misma especie. Generalmente las plantas y los microorganismos son menos sensibles a la radiación que, por ejemplo, los mamíferos. Incluso dentro de un mismo individuo, sus células tienen una sensibilidad diferente

(siendo las células que se dividen más rápidamente las más sensibles).

Existen efectos a corto y a largo plazo. Si una dosis de radiación grande es aplicada sobre el cuerpo en muy poco tiempo, los efectos biológicos que produce en un período de horas o días posteriores a la irradiación se denominan efectos a corto plazo, y se caracterizan por vómitos, fiebre, shock e incluso la muerte. Los efectos que pueden manifestarse a largo plazo se dan generalmente por exposiciones pequeñas a lo largo de los años (como el producido por diagnóstico por rayos X). Los efectos cancerígenos y genéticos son un ejemplo de estos. Los profesionales de la salud establecen límites a las cantidades que pueden administrárseles a los pacientes sin que entren en umbrales de riesgo.

2.2. Cáncer de mama

El cáncer de mama consiste en un crecimiento anormal y azaroso de las células que forman la mama. La mayoría de los tumores que se generan en la mama son benignos (no cancerosos), estos ni se extienden al resto del organismo ni son peligrosos. Se deben principalmente a factores genéticos y sus únicos efectos son inflamatorios y molestos. Los malignos pueden ser localizados o haberse extendido a través de los vasos sanguíneos o linfáticos, y se clasifican en varios tipos dependiendo del lugar de la mama donde se ha producido el crecimiento anormal y el estado en el que se encuentre:

-Carcinoma *ductal in situ*: se origina en las células de las paredes de los conductos mamarios. Es un cáncer muy localizado, que no se ha extendido a otras zonas ni ha producido metástasis. La tasa de curación en este caso es aproximadamente del 100%.

-Carcinoma *ductal infiltrante (o Invasivo)*: se origina en el conducto mamario pero consigue atravesarlo y pasar al tejido adiposo de la mama, y puede llegar a extenderse por partes del cuerpo. Es el más frecuente y se da en el 80% de los casos.

-Carcinoma *lobular in situ*: se origina en las glándulas mamarias y, a pesar de no tratarse de un verdadero cáncer, aumenta el riesgo de que la mujer pueda desarrollarlo en el futuro. Una vez detectado es importante que la mujer se realice una mamografía de control al año y varios exámenes clínicos para vigilar el posible desarrollo del cáncer.

-Carcinoma *lobular infiltrante (o Invasivo)*: se inicia en las glándulas mamarias pero es capaz de extenderse y destruir otros tejidos. Entre el 10% y el 15% de los tumores de mama son de este tipo. Este carcinoma es más difícil de detectar a través de una mamografía.

-Carcinoma *Inflamatorio*: es un cáncer poco común, representa un 1% del total de los tumores cancerosos de la mama. Es agresivo y de rápido crecimiento. La apariencia de la piel se vuelve gruesa y ahuecada, y pueden aparecer arrugas y protuberancias. Estos síntomas se deben al bloqueo que producen las células cancerosas sobre los vasos linfáticos.

En España se diagnostican cada año 15.000 nuevos casos; actualmente se manejan cifras que indican que una de cada 16 españolas tendrá cáncer de mama. En la franja de edad de 45 a 55 años, el cáncer de mama es la primera causa de muerte por cáncer en mujeres de nuestro país por delante del cáncer de pulmón. La tasa de curación en España ronda el 60%. En el caso de recaídas la estimación de supervivencia es de unos dos años de vida, lo que indica que en la realidad hay mujeres con una supervivencia corta (de menos de un año) y en otros casos esta cifra se eleva por encima de los cuatro años. La detección precoz, cuando el tumor no está extendido ni ha evolucionado, hace que el porcentaje de curación se eleve casi al 90%.

La Sociedad Americana del Cáncer (ACS) estima que cada año se diagnostican unos 190.000 nuevos casos de cáncer invasivo de mama en Estados Unidos. La supervivencia de cinco años para mujeres estadounidenses con cáncer mamario localizado es en la actualidad del 97%.

Como curiosidad, el cáncer de mama también afecta a los hombres pero en un porcentaje mucho menor (1%) comparado con la población femenina. Los síntomas son similares a los que presenta la mujer, y el pronóstico y tratamiento es el mismo. La diferencia más significativa que se puede apreciar es que se detecta un poco más tarde (ya que al darse en pocos hombres, se piensa en otros diagnósticos antes que en un cáncer de mama).

Los índices de mortalidad del cáncer de mama disminuyeron de 1992 a 1996 significativamente, probablemente debido a los mejores diagnósticos y a las mejoras en los tratamientos.

El cáncer tiene un tratamiento y pronóstico diferentes según la etapa en la que se encuentre. El Comité Conjunto Americano del Cáncer utiliza el sistema TNM:

- La letra T seguida de un número del 0 al 4. Indica el tamaño del tumor y la propagación a la piel o a la pared del tórax (debajo de la mama). Cuando más alto es el número, mayor es la propagación.

- La letra N seguida por un número del 0 al 3. Indica si el cáncer se ha propagado a los ganglios linfáticos de la mama, y si éstos están en contacto con otras estructuras.

- La letra M seguida de un número 0 o 1. Indica si se ha extendido a otros órganos.

La numeración se realiza mediante notación romana del I al IV:

- Estado I. El tumor es menor de 2 cm y no hay metástasis.

- Estado II. El tumor no mide 2 cm pero los ganglios de la axila están afectados, o mide entre 2 y 5 cm (se haya extendido o no), o mide más de 5 cm pero los ganglios de la axila no están afectados.

- Estado III. Subdividido en estado IIIA y IIIB. El IIIA indica que mide menos de 5 cm y se ha extendido por los ganglios axilares, y éstos están unidos entre ellos o a otras estructuras o que mide más de 5 cm y los ganglios axilares están también afectados. El IIIB indica que el cáncer se ha extendido a otros tejidos cercanos a la mama (como la piel o la pared torácica) o que el cáncer se ha diseminado a los ganglios dentro de la pared torácica.

- Estado IV. Indica que el cáncer se ha extendido a otras estructuras del cuerpo (huesos, pulmones...).

Las causas que generan el cáncer de mama son desconocidas, pero sí se saben algunos factores de riesgo, es decir, situaciones que aumenten las probabilidades de padecer la enfermedad. Los factores de riesgo conocidos son los siguientes:

- Sexo: mayor probabilidad en las mujeres.

- Edad: el 60% de los casos se da en mujeres de más de 60 años, el porcentaje es más elevado aún después de los 75 años.

- Genes: existen dos genes identificados que, cuando se produce algún cambio o mutación en ellos, se relacionan con una mayor probabilidad de desarrollar el cáncer de mama. Se identifican como BRCA1 y BRCA2, según algunos estudios parece que entre el 50% y el 60% de mujeres que han heredado estos genes mutados pueden desarrollar el cáncer antes de los 70 años.

- Antecedentes familiares: cuando un pariente de “primer grado” (madre, hermana o hija) ha

padecido cáncer de mama, se duplica el riesgo de padecerlo. Si lo ha sufrido un pariente “más lejano”, el aumento del riesgo es moderado.

- Antecedentes personales: una enfermedad mamaria benigna previa aumenta el riesgo en las mujeres que tienen un elevado número de conductos mamarios, y el riesgo de padecer cáncer en la otra mama en las mujeres que ya han sufrido uno es distinto a la reaparición del primero.

- Raza: las mujeres blancas son más propensas a tener esta enfermedad, pero las de raza negra tienen una mortalidad mayor (probablemente debido a que se les detecta en estados más avanzados). Las que tienen un riesgo más bajo son las asiáticas e hispanas.

- Períodos menstruales: cuanto antes se comienza con la menstruación, mayor es el riesgo (de dos a cuatro veces mayor) si se compara con aquellas que comenzaron más tarde. Lo mismo ocurre con la menopausia: las mujeres con una menopausia tardía tienen mayor riesgo. El embarazo después de los 30 años también aumenta el riesgo. Estos factores, aunque muy frecuentes, suelen tener poca incidencia sobre el riesgo de padecer cáncer.

La mama está formada por tejido glandular y graso. El glandular es relativamente denso y en la radiografía tiene valores elevados de gris/blanco. El tejido graso es menos denso y aparece en la mamografía con valores pequeños de gris/negro. Los tumores tienen una densidad similar a la del tejido glandular y se muestran con la misma tonalidad que éste. En algunas ocasiones, se puede observar también un pequeño moteado sobre la imagen que corresponde a depósitos de calcio (calcificaciones) de tamaño minúsculo.

2.3. Estudio de la densidad de mama

El estudio de la densidad mamaria ha sido abordado con especial interés durante las últimas décadas. Las principales razones que justifican este interés pueden resumirse en que la densidad mamaria es un indicador del riesgo de padecer cáncer de mama, y es determinante en la sensibilidad del análisis de la mamografía. Como ya se ha comentado anteriormente, y al igual que ocurre con otros tipos de tumores, las posibilidades de curación del cáncer de mama aumentan espectacularmente si éste se detecta en sus primeras fases, lo que confirma que el mejor método para combatir esta enfermedad es el diagnóstico precoz. La confirmación diagnóstica se hace generalmente con mamografías. Actualmente los programas de cribado mamográfico tienen como objetivo el diagnóstico del cáncer de mama en una fase precoz, antes de la aparición de los síntomas. Aproximadamente el 80-85 % de los cánceres pueden ser detectados con mamografía, sin embargo, en aquellos casos en los que la mama tiene una mayor densidad el diagnóstico resulta más complicado, ya que en este caso el tejido denso tiene una absorción frente a los rayos X similar a la de algunos tumores. Ello provoca que existan tumores no palpables en mamas densas que pueden permanecer ocultos hasta que su desarrollo alcanza la fase clínica y su tamaño es mayor que el deseado para considerarlo como diagnóstico precoz, lo cual conlleva a un tratamiento con mayor morbilidad y peor pronóstico que los tumores detectados en fases precoces y pequeños tamaños [1-5].

Además de la dificultad que entraña la detección de determinados tumores, existen estudios que demuestran un aumento del riesgo de desarrollar cáncer de mama en las mujeres con mamas densas, el cual puede ser de 4 a 6 veces superior al de las mujeres en las que predomina el tejido graso. Existe sin embargo cierta controversia acerca de si la asociación densidad-cáncer se debe a un riesgo intrínseco de desarrollar la enfermedad o únicamente a la dificultad de identificar los tumores incipientes debido al problema de enmascaramiento de rayos X citado. De un modo u otro es un hecho demostrado que las mujeres con mamas densas tienen una mayor probabilidad de desarrollar cáncer, lo que ratifica el interés de conocer la densidad

[6-8].

Debido a los inconvenientes que presenta una mama densa en relación al desarrollo de un tumor, desde hace ya varias décadas se vienen realizando diversos estudios que tratan de clasificar la mamografía según su densidad con el objetivo de establecer un factor de riesgo frente al cáncer de mama. Esta medida puede ser de gran utilidad por diversos motivos. En primer lugar, este factor de riesgo podría influir en la elección de un protocolo de cribado distinto del habitual, como puede ser realizar mamografías con mayor periodicidad, o utilizar otros test de imagen médica como la resonancia magnética o la ecografía. En segundo lugar podría utilizarse para establecer la necesidad de un análisis más cuidadoso de la mamografía, como por ejemplo la doble lectura o la lectura asistida por ordenador (CAD). Finalmente, de manera reciente se ha propuesto la utilización de la densidad mamográfica como fenotipo intermedio en la investigación de cáncer de mama, lo que supone que dicha característica (evaluable en los propios programas de cribado) podría ser utilizada como marcador de riesgo de forma similar a la utilización de la hipertensión arterial en el estudio de la enfermedades cardiovasculares.

Desde que en la década de los 70 se propusiera por primera vez un método de clasificación de la mama en función de su densidad, han ido apareciendo diversas propuestas con el mismo objetivo. Algunos trabajos proponen fundamentalmente clasificar las mamografías según la morfología que adquiere el tejido fibroglandular o TFG (métodos cualitativos) mientras que otros lo hacen determinando la cantidad de TFG en relación al volumen de la mama (métodos cuantitativos). Los métodos cualitativos clasifican preferentemente por la morfología pero no cuantifican la cantidad de tejido denso. Los métodos cuantitativos, por contra, clasifican fundamentalmente en base a la relación porcentual que existe entre tejidos graso y fibroglandular, atendiendo a esta única razón, o añadiendo una descripción morfológica del TFG. En la aplicación de estos métodos puede decirse que con excepción del descrito por Boyd, que sería estrictamente cuantitativo, el resto son semicuantitativos.

En 1976, Wolfe propuso una clasificación que utilizaba cinco categorías para establecer la densidad. Se trataba de un método cualitativo para establecer la densidad del TFG partiendo de la morfología del mismo (densidad asociada a estructuras ductales). Este método fue utilizado durante mucho tiempo y mostraba el riesgo para desarrollar cáncer de mama asociado a cada uno de los patrones que describía: mamografía casi completamente grasa con trabéculas (N1), formaciones ductales que ocupan hasta un 25% del volumen (P1), patrón ductal prominente que ocupa más del 25% (P2), displasia mamaria severa (DY) y quasy-DY (QDY).

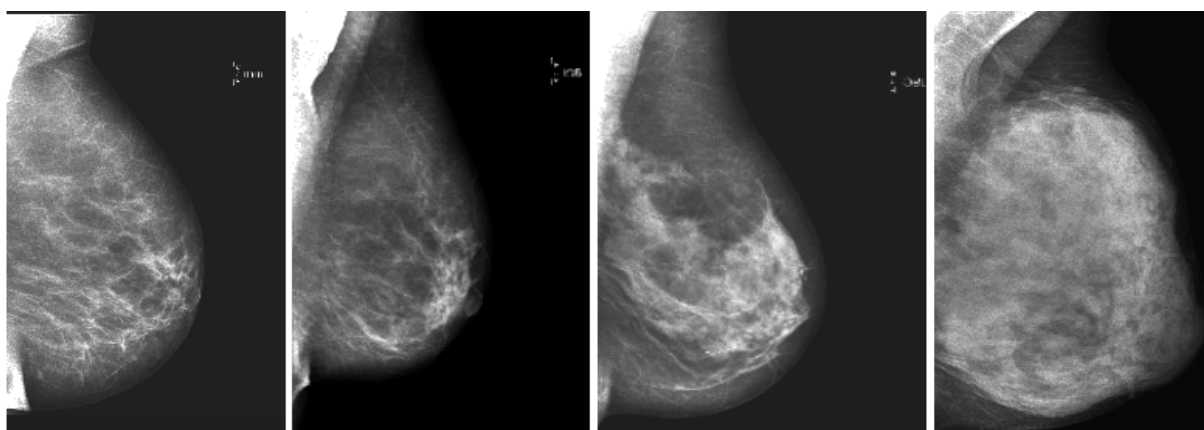


Imagen 9. De izquierda a derecha, mama grasa con imágenes lineales finas y densas que corresponden a fibras (N1); mama con predominio graso e imágenes tubulares de conductos galactoforos (P1); mama con imágenes tubulares de conductos galactoforos (P2); y mama densa

(cuya densidad borra total o parcialmente las estructuras tubulares).

El método de Tabar establece cinco patrones mamarios, pero en este caso basados en una correlación de la histología (nivel microscópico) con la mamografía (nivel macroscópico).

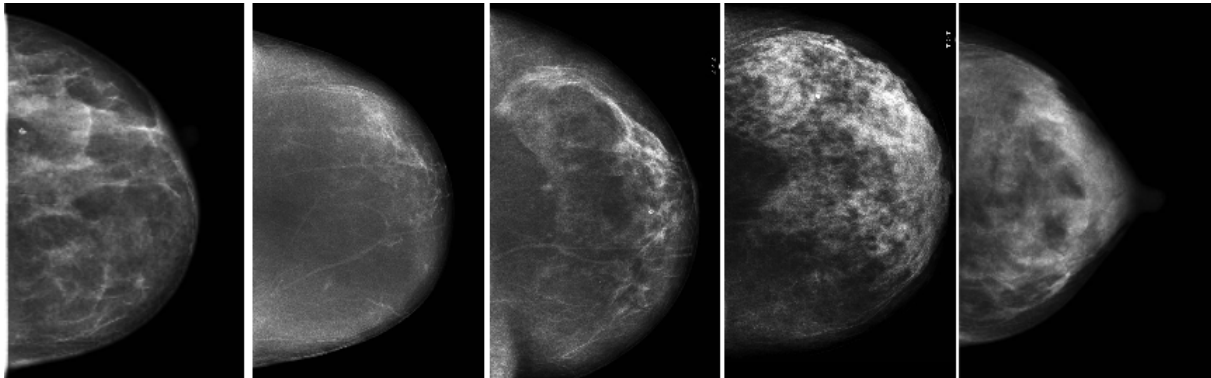


Imagen 10. de izquierda a derecha, mujer joven con densidades nodulares pequeñas, tejido adiposo y fibroso y ligamentos de Cooper (Patrón I); mujer mayor con predominio graso (Patrón II); mujer mayor con predominio graso con patrón ductal prominente retroareolar (Patrón III); patrón nodular grande que no cambia con el tiempo (Patrón IV); y tejido muy denso que borra todas las estructuras anatómicas (Patrón V).

Más adelante se establecieron sistemas de clasificación que cuantificaban únicamente el porcentaje de tejido denso y no su forma, lo que resultaba menos ambiguo. En este sentido Boyd propuso seis niveles de densidad sin considerar la apariencia del TFG, estableciendo de modo cuantitativo los porcentajes de tejido en relación al volumen total de la mama (0 %, [1-10 %], [11-25 %], [26-50 %], [51-75 %] y [76-100 %]).

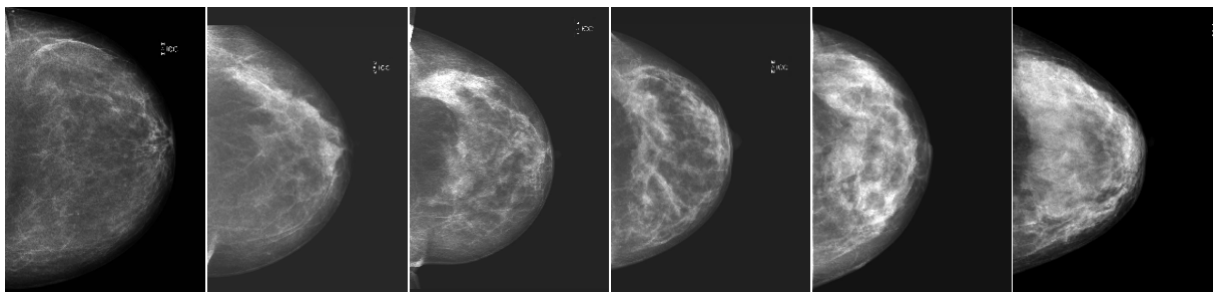


Imagen 11. de izquierda a derecha, 0% de tejido glandular (A); 0% a 10% de tejido glandular (B); 10% a 25% de tejido glandular (C); 25% a 50% de tejido glandular (D); 50% a 75% de tejido glandular (E); y 75% a 100% de tejido glandular (F).

Finalmente el método propuesto por la American College of Radiology (ACR), conocido como BI-RADS (Breast Imaging Reporting and Data System), propone una clasificación cuantitativa, con cuatro tipos de mama definidos por los umbrales de 25 %, 50 % y 75 % de densidad de TFG. Este es el método que tiene mayor aceptación [9,10].

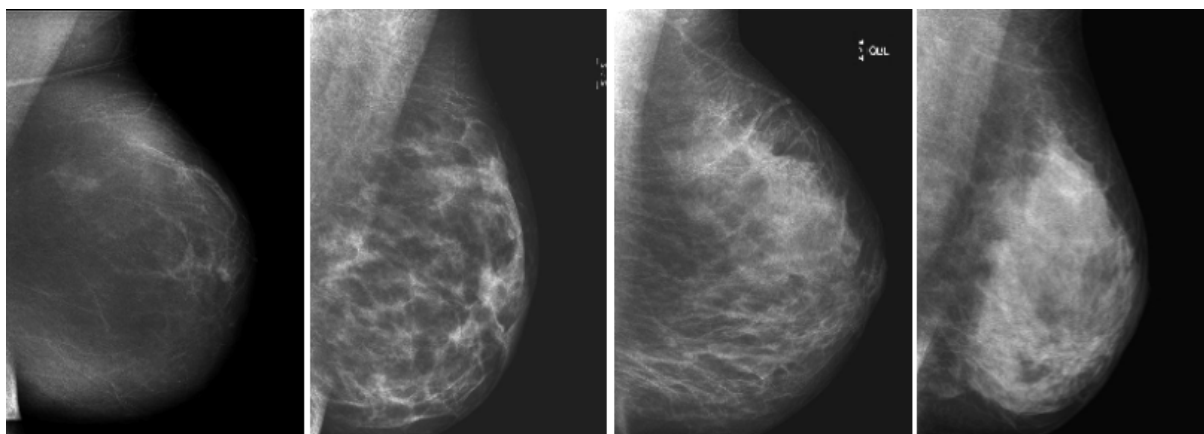


Imagen 12. de izquierda a derecha, menos del 25% de TFG, casi toda grasa (Patrón 1); aprox. 25% al 50% de TFG, densidades dispersas o parcheadas de TFG (Patrón 2); aprox. 51% al 70% de TFG, tejido heterogéneamente denso (Patrón 3); más del 75% de TFG, muy denso (Patrón 4).

Independientemente del método empleado, uno de los mayores inconvenientes que conlleva la clasificación de la mama mediante inspección visual radica en la subjetividad con que el lector de mamografías clasifica éstas en una u otra categoría. Este proceso requiere de una gran experiencia en lectura y, a pesar de todo, existe gran variabilidad en la clasificación de las mamografías.

Los problemas de inconsistencia se deben tanto a cuestiones de subjetividad en la interpretación de la mamografía como a una percepción errónea del criterio de clasificación empleado. Dos alternativas, no necesariamente excluyentes, podrían emplearse para tratar de minimizar este problema. En primer lugar, si se dispusiera de un amplio conjunto de mamografías clasificadas por distintos expertos, resultaría interesante desarrollar los algoritmos necesarios que permitieran recuperar de dicha base de datos una o varias imágenes similares a la que está siendo analizada. Con ello el radiólogo podría conocer el diagnóstico que distintos expertos han realizado sobre imágenes con características similares. Actualmente existen algunas bases de datos públicas de mamografías como la *Mammographic Images Analysis Society Digital Mammogram Database* (MIAS) o la *Digital Database for Screening Mammography* (DDSM), pero no incluyen ningún motor de búsqueda como el propuesto y las imágenes están etiquetadas en base al criterio de un único experto. En segundo lugar pueden emplearse técnicas basadas en reconocimiento de formas y visión por computador para calcular de forma automática y objetiva la cantidad de tejido denso presente en la mamografía.

2.4. Clasificación asistida por ordenador

El reconocimiento de formas (RF) es una disciplina que estudia cómo construir sistemas automáticos, capaces de emular aspectos perceptivos propios del comportamiento humano. El término reconocer equivale a clasificar en C clases o categorías los distintos objetos observados. Por ejemplo, dada una mamografía podría pedirse a un sistema de RF que la clasifique en uno de las cuatro categorías de densidades establecidas por BI-RADS.

La llamada aproximación inductiva al RF requiere del uso de un conjunto de muestras de aprendizaje, que son utilizadas para crear un modelo. Posteriormente, un clasificador usará este modelo para identificar (clasificar) objetos no vistos durante la etapa de aprendizaje. La aproximación inductiva lleva asociada implícitamente el concepto de aprendizaje.

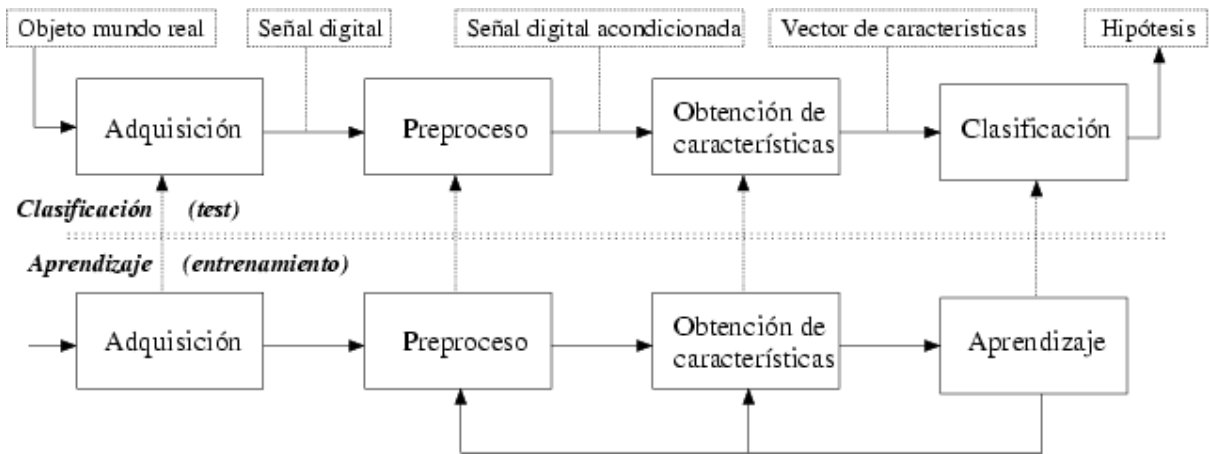


Imagen 13. Esquema de un sistema de reconocimiento de formas.

El sistema puede trabajar en modo aprendizaje o en modo clasificación. En ambos casos se comienza con la adquisición de un objeto del mundo real (por ejemplo, captura de una imagen). A continuación se realiza un preproceso cuyo objetivo es realzar y acondicionar los objetos de interés. Seguidamente, en la etapa de obtención de características se realizan ciertas medidas sobre el objeto que se desea analizar. La idea es obtener una serie de valores numéricos o símbolos (vector de características) representativos del objeto. Esta etapa es crucial, ya que dependiendo del tipo de características con las que trabajemos, será más o menos viable discriminar entre objetos de distintas clases. Finalmente, dependiendo de si se trabaja en modo aprendizaje o en modo clasificación, estas características se utilizarán para aprender (entrenar) un modelo o para clasificar el objeto observado en una de las C posibles clases, teniendo en cuenta el modelo aprendido.

Existen distintas aproximaciones a la hora de diseñar el modelo que representa los objetos de nuestro problema. En la aproximación estadística, los objetos se representan en forma de n características o medidas, conocidas como vector de características. Estos vectores pueden verse como un punto en un espacio de n dimensiones. El objetivo es escoger aquellas características que permitan distribuir los vectores de características (puntos) de manera que las distintas clases o categorías ocupen regiones en el espacio n -dimensional lo más compactas y disjuntas posible. Con ello, se establecen unas fronteras de decisión en el espacio de características que tratan de separar los vectores de cada clase. Estas fronteras de decisión se determinan a partir de las distribuciones de probabilidad de los vectores de características de cada clase, obtenidos de las muestras de aprendizaje. Durante la etapa de clasificación, ante una muestra desconocida (muestra de test), ésta se clasificará en una clase u otra en función de la posición que ocupe respecto a las fronteras de decisión.

3. Clasificación de la densidad de mama asistida por ordenador

Existen multitud de técnicas estadísticas para definir las fronteras de decisión y realizar la clasificación. Las prestaciones del sistema de RF dependerán, entre otros factores, del preproceso, del clasificador diseñado y del tipo de características empleadas. A continuación se realizará un resumen del tema relacionado con la segmentación de la radiografía en la parte interesante (la mama), una breve reseña de los clasificadores más empleados en la clasificación de la densidad y un resumen del estado del arte en que se encuentra actualmente esta clase de estudios.

3.1. Segmentación de mama

En una situación ideal, la mamografía sólo contendría la información relevante de la mama para su estudio. En la realidad, sin embargo, antes de realizar cualquier clase de experimentación informática con las mamografías hay que someter las imágenes a un preproceso para eliminar la información innecesaria. En el caso concreto de las mamografías la información innecesaria la forma el fondo de la imagen y el músculo pectoral.

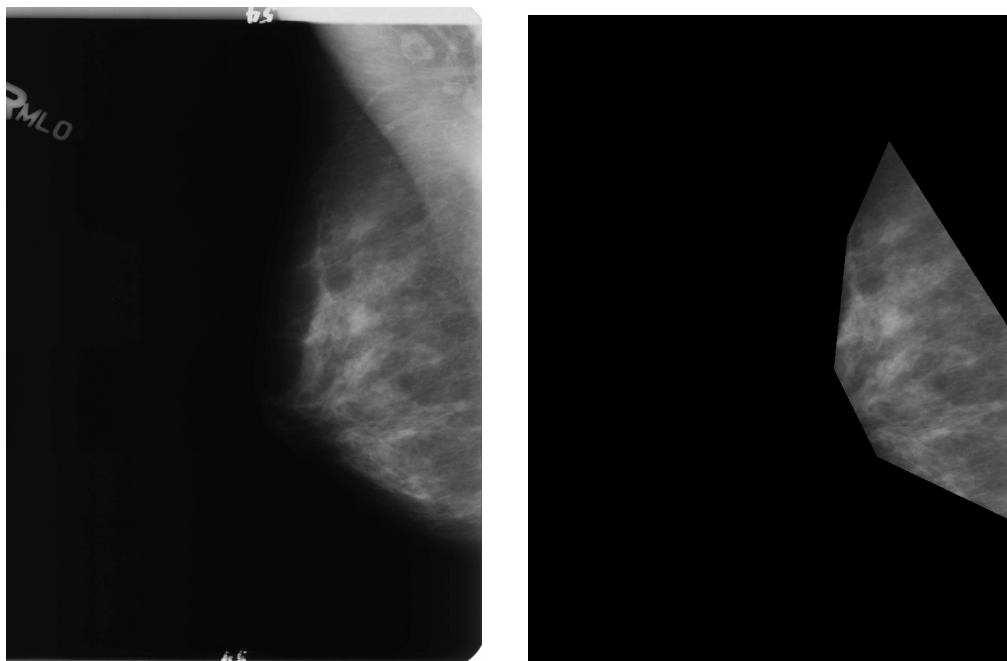


Imagen 14. Mamografía original y mamografía tras ser preprocesada.

En el fondo de la imagen puede haber bordes luminosos o rótulos que pueden entorpecer el estudio de los niveles de gris, mientras que el músculo pectoral aparece en las mamografías con tonos de gris muy elevados (Imagen 14) que causan los mismos problemas. También se utiliza la segmentación para descartar los píxeles negros del fondo, ya que son inútiles para el estudio y sólo añadirían más tiempo de cálculo al proceso.

La opción más simple (y costosa) es realizar una segmentación manual de las mamografías. Se obtiene exactamente la mama, descartando el resto de información. Sin embargo, no es un sistema viable si se dispone de una gran cantidad de imágenes.

La segmentación de mamografías automática es un problema abierto en la comunidad científica. A pesar de los métodos propuestos, no existe un procedimiento que permita

obtener de forma automática el 100 % de la mama, descartando completamente la información del músculo pectoral y del fondo de la mamografía.

Para separar la mama del fondo, se han propuesto varios métodos de detección de objetos. En un primer momento se binariza la imagen, quedando los “objetos” en blanco y el fondo en negro. Después, se utiliza sólo el objeto de mayor área, que será la mamografía. El problema principal de esta aproximación es que si la imagen presenta bordes luminosos, se incluirán en el objeto de mayor área. Otro problema es que el valor de gris que se utilice para umbralizar la imagen afectará al límite entre lo que se considere fondo y lo que se considere objeto, de manera que el segmentador puede quedarse con menos mama de la necesaria o pasar información de fondo innecesaria como parte de la mama. Estas complicaciones pueden atenuarse realizando un preumbralizado con los valores de gris más elevados y más bajos de la mamografía, eliminando los posibles bordes luminosos y la incierta frontera entre objetos y fondo.

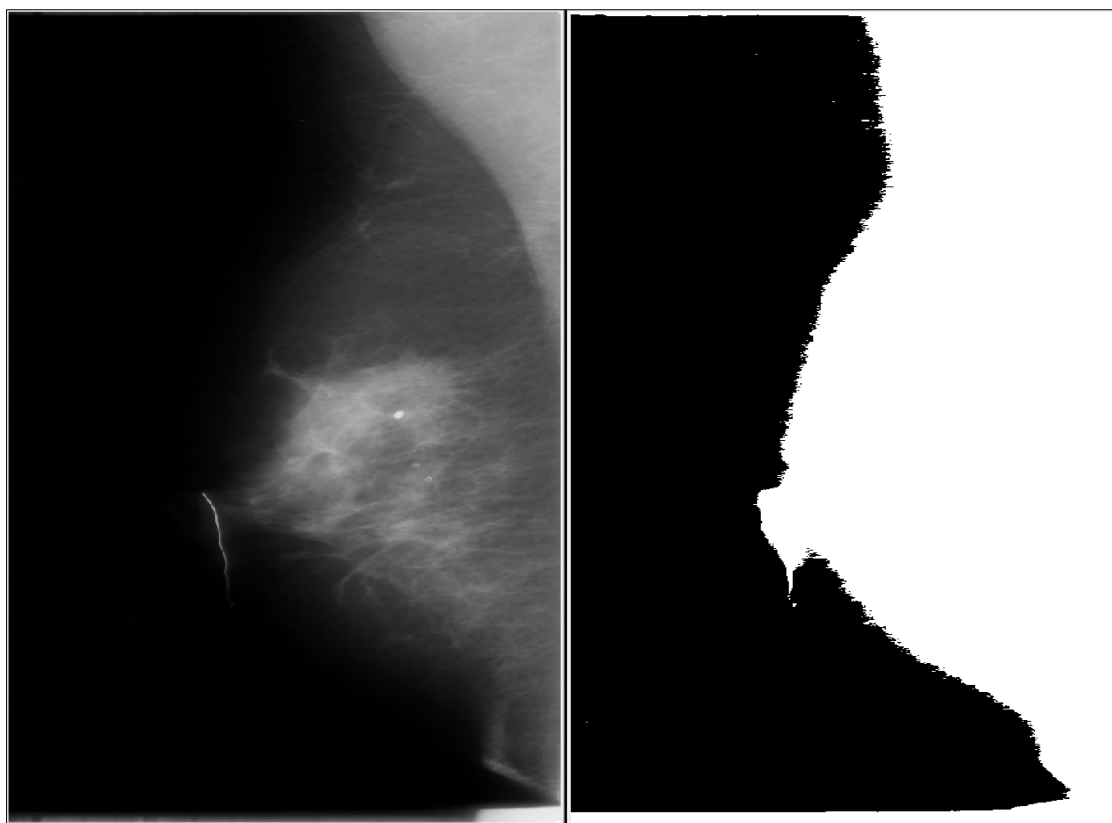


Imagen 15. Umbralización en la que se aprecian los bordes blancos superiores e inferiores que se incluirían dentro de lo que se considera mama si no se eliminan.

Otros autores optan por métodos relacionados con segmentación basada en semillas. Se define una rejilla que divide la imagen en zonas independientes (denominadas ventana). De cada ventana se debe encontrar el píxel de mayor brillo (máximo local) que es lo que se llama semilla.

Dependiendo del tamaño de ventana, se obtendrán un número más o menos elevado de semillas que representan a toda la población de píxeles. Cada zona al ser independiente se puede estudiar en paralelo. La idea fundamental del algoritmo de segmentación es la asignación de píxeles, con el nivel de gris más o menos parecido, a la misma región. Una

región se puede definir como el conjunto de píxeles, vecinos entre sí, cuya diferencia de tonalidad con la semilla de la región pertenece al mismo rango. Una vez “plantadas” las semillas, se crean estas regiones que marcan las zonas de riesgo en la mamografía. La fiabilidad del proceso depende, en todo momento, del plantado de semillas y de la precisión con la que se realice su posterior colocación. El amplio abanico de opciones e implementaciones posibles (*kdtrees*, árboles binarios balanceados, colas de vecinos, threads, soluciones *ad-hoc*...), junto con los diferentes conjuntos de test empleados hace difícil determinar qué opciones son las que obtienen mejores resultados.

Algunos autores han intentado segmentar la imagen del fondo utilizando matrices de coocurrencias y detección de texturas. La matriz de coocurrencia representa la dependencia espacial entre dos niveles de gris de la imagen. Esta matriz se forma para cada vector dado, definiendo la separación espacial entre dos píxeles sobre los cuales se calcula la dependencia de niveles de gris, reagrupando las estimaciones de la densidad de probabilidad conjunta de todas las parejas de puntos separados por el vector. Al final, todos los componentes son divididos por el número total de parejas analizadas. Se obtienen varios parámetros de evolución a partir de la matriz, y con ellos se obtiene la textura correspondiente. Los parámetros más significativos son: el contraste, que mide la repartición de los coeficientes de la matriz con respecto a la diagonal principal; el momento angular de segundo orden, que indica el nivel de homogeneidad de la textura; y el momento de orden cuatro, que establece relaciones entre tres píxeles. Sin embargo, los resultados obtenidos por estos procedimientos dependen en gran medida de la calidad de la mamografía (en las muy claras, por ejemplo, suele tomar como parte de la textura de la mama zonas claras de la imagen que en realidad no corresponden a la mama).

Una vez se ha conseguido separar la mama del fondo, eliminando posibles errores gráficos en la mamografía, hay que eliminar el músculo pectoral que, al igual que el fondo, no tiene ningún valor a la hora de clasificar la mama y puede producir errores (a parte de aumentar el tiempo de procesamiento, claro). Esta segmentación es un proceso aún más delicado que el de separar la mama del fondo. Sea cual sea la estrategia a seguir, el primer paso es decidir la orientación de la mama (si es izquierda o derecha) para saber en qué lugar hay que empezar a buscar el músculo pectoral. Afortunadamente, es un proceso bastante sencillo y con múltiples aproximaciones: desde analizar la curvatura de los dos bordes de la mama hasta calcular el centro de masa del objeto y ver en qué mitad hay más píxeles, pasando por la división de la imagen en regiones y estudiándolas individualmente. El problema de verdad es, una vez determinada la orientación, conseguir segmentar de manera adecuada el músculo pectoral.

Una aproximación rápida y no demasiado acertada es suponer unas inclinaciones predeterminadas para el músculo pectoral, segmentar la mama teniendo en cuenta cada una de las inclinaciones y ver cual de las imágenes resultantes es la más adecuada (comparando la parte segmentada que representa el músculo, por ejemplo). Aunque es fácilmente implementable, se desaconseja esta aproximación a menos que se necesite una segmentación aproximada y no demasiado acertada.

Entre las técnicas de separación del músculo pectoral, hay autores que se decantan por el crecimiento de regiones. Para ello, se planta una semilla en el músculo pectoral (que, una vez determinada la orientación de la mama, será el primer píxel del lado recto) y se utiliza un algoritmo de crecimiento de regiones para determinar la zona que pertenece al músculo. Existen multitud de algoritmos de crecimiento de regiones así como una elevada cantidad de métodos correctores y limitadores para impedir que se incluyan zonas de la mama en la región correspondiente al músculo. Una vez más, el elevado número de opciones y

conjuntos de tests impide obtener una estimación de qué combinación puede ser la mejor para obtener el segmentado más certero.

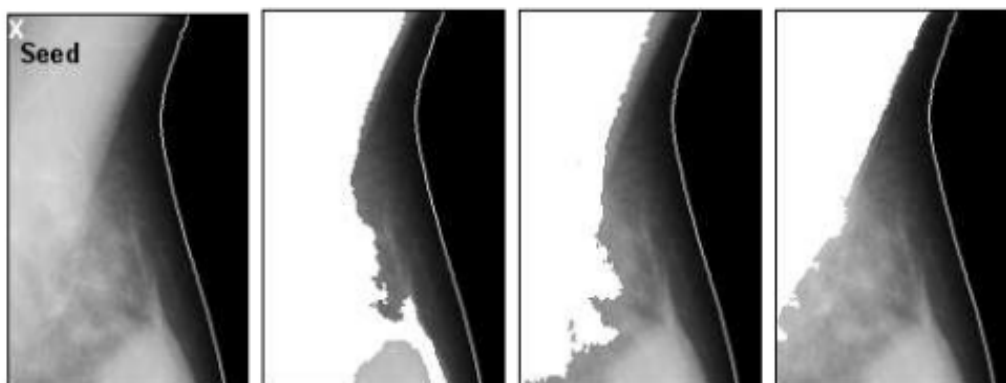


Imagen 16. Crecimiento de regiones para eliminar el músculo pectoral.

Otros métodos se basan en el uso de la geometría y la determinación de rectas para eliminar el músculo, al igual que hemos comentado anteriormente, pero empleando técnicas más complejas. Transformar la imagen mediante la transformada de Hough para detectar las líneas rectas e intentar elegir la que representa el límite entre el músculo y la mama o usar el mismo principio geométrico citado anteriormente pero usando ROIs (Regions Of Interest, ventanas concretas de la imagen) y métodos iterativos son algunas de las opciones que se están barajando actualmente.

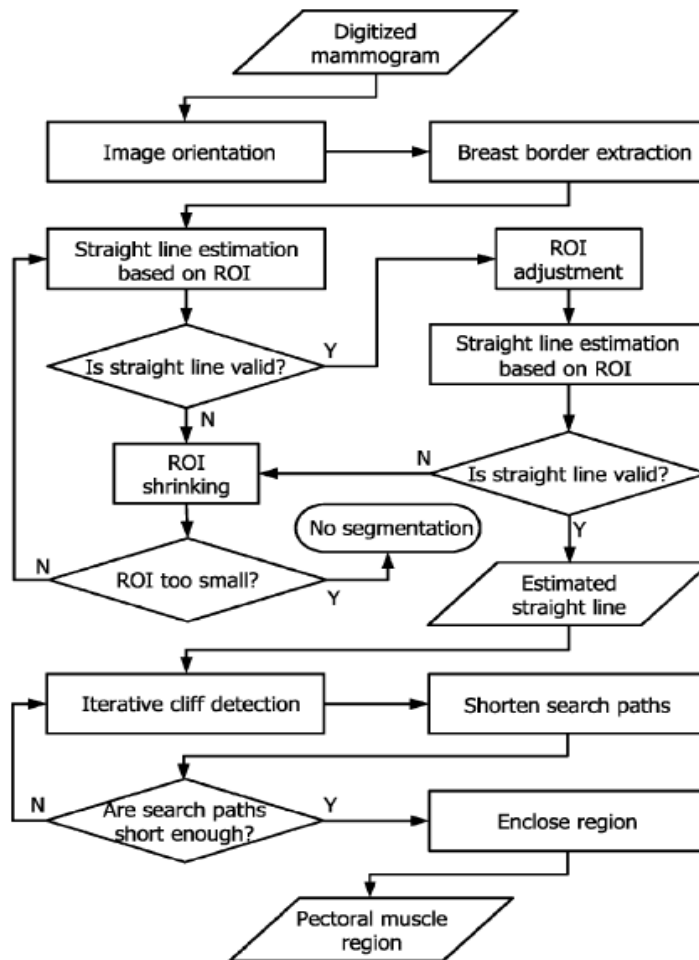


Imagen 17. Método iterativo presentado por Sze Man Kwok, en el que se marca inicialmente una línea recta como límite del músculo y se corrige en cada iteración.

Por lo tanto, para abordar cualquier estudio asistido por ordenador de un conjunto de mamografías, o se tiene la suerte de disponer de una base de datos correctamente segmentada o hay que decidir que método se seguirá para preparar las imágenes, separándolas del fondo y eliminando el músculo pectoral.

3.2. Clasificadores

Redes Neuronales

Basadas en los sistemas nerviosos, las Redes Neuronales son redes de interconexión de neuronas que ofrecen una salida a partir de unos datos de entrada. Cada neurona recibe unas entradas y produce una salida, que se obtiene a partir de una función de propagación (normalmente un sumatorio), una función de activación (opcional, que modifica el resultado de la anterior) y una función de transferencia (que convierte la salida a un rango de números concreto).

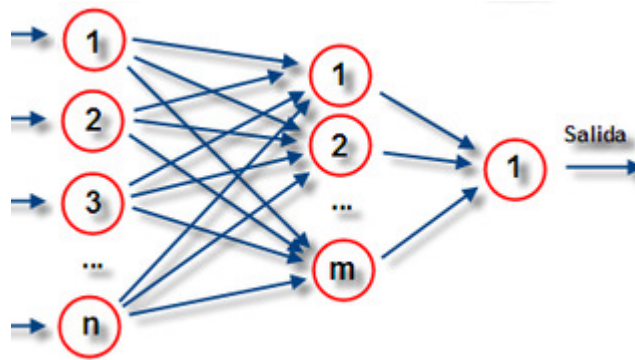


Imagen 18. Red Neuronal con n neuronas de entrada, m intermedias y una de salida.

Las ventajas principales de las Redes Neuronales Artificiales son:

- Aprendizaje adaptativo: las Redes Neuronales son sistemas didácticos autoadaptativos, gracias a la capacidad de autoajuste de las neuronas que componen el sistema. Durante el aprendizaje, los enlaces de las neuronas se ajustan para obtener resultados específicos. Y también hay Redes que continúan aprendiendo durante toda su “vida”, y no paran en la fase de aprendizaje.

- Auto-organización: usando su capacidad de aprendizaje adaptativo para organizar la información que reciben durante el aprendizaje o la operación en concreto. Esta auto-organización provee de generalización a la Red, lo que le permite responder apropiadamente cuando se le presentan datos o situaciones a las que no había sido expuesta anteriormente.

- Tolerancia a fallos: en las Redes Neuronales, si se produce un fallo en un pequeño número de neuronas el sistema no sufre una caída (aunque el comportamiento del mismo sí que se verá afectado). La principal razón se debe a que la información se encuentra distribuida en las conexiones entre neuronas, existiendo un cierto grado de redundancia.

- Operación en tiempo real: es posible gracias a la implementación paralela de la Red

Support Vector Machines (SVM)

Son un conjunto de métodos de aprendizaje supervisados que pueden usarse para clasificación y regresión. Originalmente se usa sobre problemas de dos clases, en los que se divide la información de entrada en dos conjuntos de vectores en un espacio n -dimensional. El SVM construirá un hiperplano que separe el espacio maximizando el margen entre los dos conjuntos de datos. Para hacerlo, construye dos hiperplanos paralelos (uno a cada lado del hiperplano separador) y los va ajustando a los conjuntos de datos.

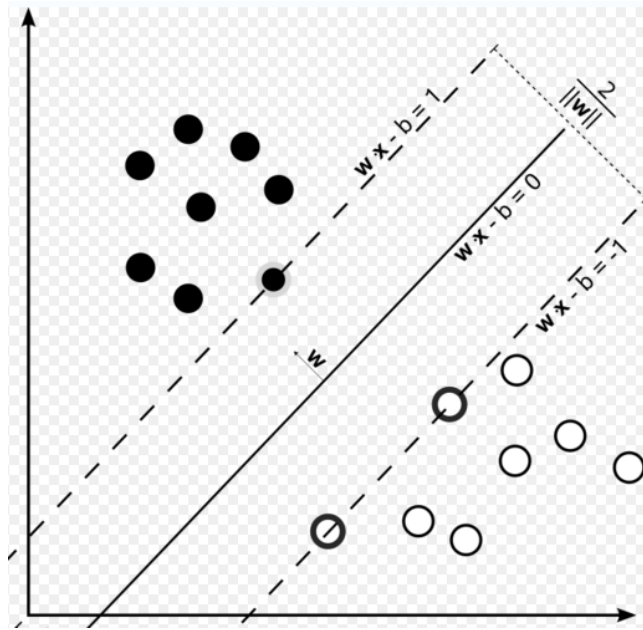


Imagen 19. Hiperplano que separa las dos clases de forma aproximadamente equidistante

Posteriormente, se han empezado a usar SVM multiclase. En estos casos el problema se subdivide en problemas de dos clases y se sigue el método descrito anteriormente para cada par de problemas.

K-vecinos

Si se considera un espacio de representación bidimensional y una serie de prototipos de una misma clase representados en él, dado un patrón cualquiera X al considerar los k prototipos más próximos a X , éstos estarán localizados en un círculo centrado en X .

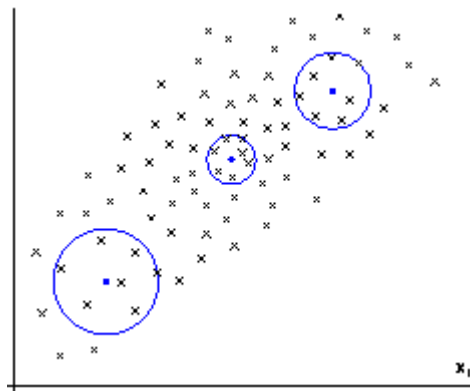


Imagen 20. Círculos con los 7 vecinos más cercanos.

El área del círculo que encierra un número fijo de puntos, k , es menor en regiones densamente pobladas que en regiones donde los puntos están más dispersos. Este sencillo planteamiento es la base de la estimación mediante los k vecinos más próximos. En espacios multidimensionales, el círculo se convierte en una *hiperesfera*, y el planteamiento anterior se puede extender fácilmente ya que el volumen de la hiperesfera que encierra k puntos está relacionado con el valor de la función de densidad de probabilidad en el centro de la hiperesfera.

De esta forma, una clase se clasificará en función de las clases de sus k-vecinos más cercanos. Este clasificador tan simple ha dado y sigue dando muchos resultados satisfactorios en muchos sistemas de decisión informáticos.

Árboles de decisión

Originalmente, los árboles de decisión eran usados en las empresas para organizar la información y las decisiones a tomar de forma secuencial, ayudando a las empresas a tomar la decisión más apropiada según los resultados previsibles.

Utilizado como método predictivo, se basa en el uso de diagramas lógicos mediante los que se representa y califica una serie de condiciones para resolver problemas. Las entradas para el árbol se presentan como un conjunto de atributos y dan como resultado una solución tomada a partir de los mismos. Los datos van pasando por los nodos que forman el árbol, según las decisiones que se tomen en ellos hasta llegar a los nodos finales, que representan el conjunto de soluciones posibles.

Por lo tanto, los árboles de decisión se utilizan para descubrir pautas en los datos, estas pautas se recogen y se organizan en modelos que se utilizarán posteriormente para hacer predicciones. Los modelos pueden ser descritos como gráficos o árboles. Cada nodo es un camino elegido sobre la base de las entradas, hasta que al final de una "hoja" un nodo es alcanzado. La "hoja" nodo representa una decisión y es usado como el resultado previsto para nuevos y desconocidos datos venideros.

Clustering

Este método no supervisado se basa en encontrar "estructuras" en un conjunto de datos no etiquetados, permitiendo la identificación de agrupaciones intrínsecas de los datos no etiquetados. Con esto, se puede realizar un proceso de etiquetado de los datos a partir de los propios datos y sin ningún tipo de información adicional.

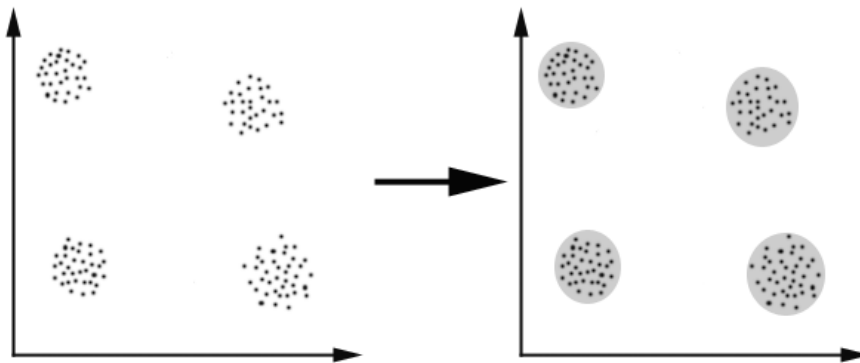


Imagen 21. Sencillo caso en el que se encuentran fácilmente cuatro "estructuras" en el conjunto de datos no etiquetados.

La variante de *clustering* en *k-means* consiste en, conocido el número de clases k que se pretende establecer, elegir de forma aleatoria un vector del conjunto de datos que representará a cada clase k (en este contexto, el vector se llamará centroide). Sucesivamente, los datos pasan de una clase k a otra en función de la distancia entre ellos y el centroide de cada clase. Cada vez que se produce un cambio entre elementos, los k centroides se recalculan.

3.3. Estado del arte

El análisis de la densidad de la mama asistido por ordenador podría ayudar a clasificar las mamografías de un modo más objetivo y coherente y, consecuentemente, establecer el factor de riesgo para desarrollar cáncer de mama con mayor precisión. Sin embargo, a pesar del indudable interés que comporta la estimación de la densidad de la mama de un modo objetivo, han sido pocos los esfuerzos realizados en este sentido dentro del campo del procesado de imágenes y análisis por ordenador.

Unos de los primeros en sugerir el uso del ordenador como herramienta de ayuda para el análisis de la densidad mamaria (1994) fueron *Boyd et al.* [7] Presentaron un método semiautomático para estimar la densidad de la mama, basado en la selección manual de dos umbrales para segmentar la mama y el tejido denso respectivamente. En este caso no se dota a la máquina de ninguna capacidad de decisión y es el experto humano quien establece el umbral que discrimina entre tejido graso y denso. El especialista empieza seleccionando un nivel de gris (*iedge*) que se usa como umbral para separar la mama del fondo. Una vez segmentada la mama a partir del *iedge*, el especialista selecciona un nuevo nivel de gris (*iDY*) que se usa como indicador del tejido denso de la mama. Los píxeles cuyo valor sea igual o superior a *iDY* se consideran tejido denso. Esta selección de píxeles se repite para cada una de las mamografías. Con los porcentajes de tejido denso y tejido no denso, se clasifica la imagen como perteneciente a una clase determinada.

El estudio original comparaba los resultados obtenidos por dos radiólogos, un médico y un observador no relacionado con la medicina, usando un conjunto privado de 60 mamografías y utilizando 6 clases de densidad. Lo más sorprendente del asunto es que el usuario que no estaba especializado en medicina demostró obtener un alto grado de concordancia con el resto de usuarios, mostrando que no era necesario personal especializado para el uso de la aplicación.

	Radiologist R1	Radiologist R2	Experienced clinician R3	Naive observer R4
R1	0.93	0.90	0.87	0.94
R2	—	0.94	0.84	0.91
R3	—	—	0.97	0.91
R4	—	—	—	0.99

Imagen 22. Tabla de correlación entre los usuarios de los experimentos.

Este método ha ganado importancia con el tiempo y ahora es conocido como *Cumulus*. Sin embargo, la necesidad de tener a una persona realizando elecciones de píxeles para cada mamografía limita el uso del ordenador a ser una interfaz en la que el usuario observa y analiza cada mamografía.

Karssemeijer et al. (1997) [11] presentan un método automático que emplea características extraídas del histograma de niveles de gris. Extrayendo regiones de interés de la mama basadas en la distancia que hay de la región al músculo pectoral, obtiene sus histogramas. Consiste en segmentar la mama en 5 partes diferentes (a un 20%, 40%, 70% y 90% del borde de la mama, respectivamente), y usar los histogramas de las 3 partes centrales (la del 40%, el 70% y el 90%) como vector de características.

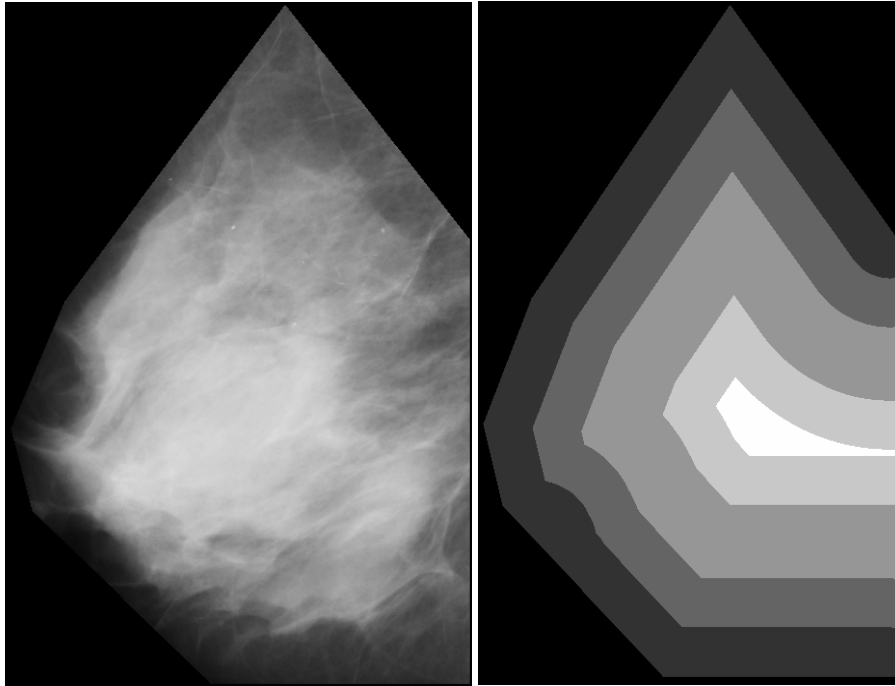
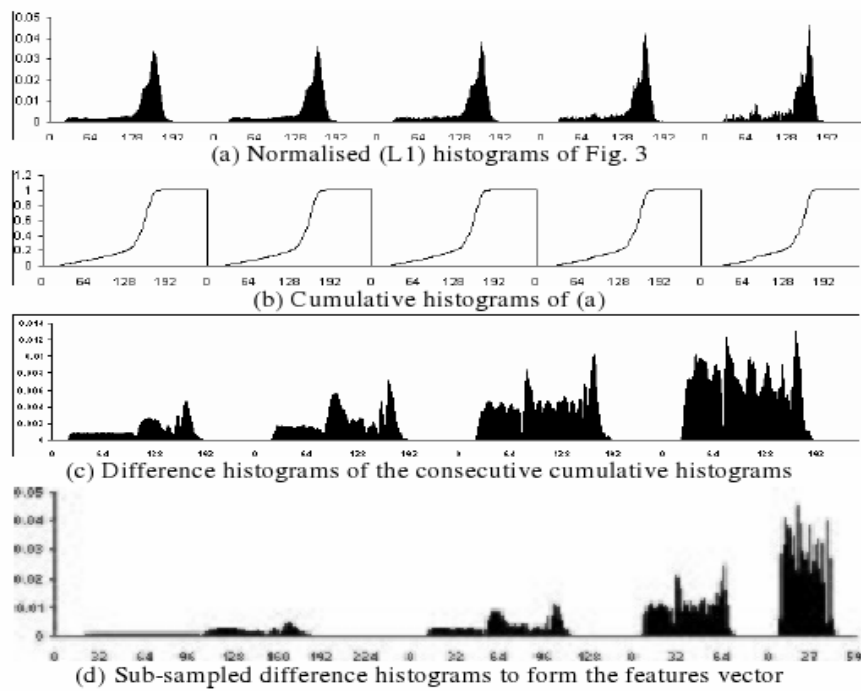


Imagen 23. *Mama segmentada en 5 partes a 20%, 40%, 70% y 90% de los bordes.*

Para la clasificación se aplica la técnica de k-Vecinos sobre el vector de histogramas resultante. Emplean cuatro clases de densidad y obtienen una concordancia del 67 % entre el sistema automático y el experto humano, utilizando una base de datos privada con 615 mamografías.

Oliver et al. [13] aplican un método que utiliza descriptores de textura como aproximación a la extracción de características y Árboles de Decisión para la clasificación en tres categorías. Consiguen una concordancia del 73 %.

En 2004, *Hadjidemetriou et al.* describen un método que emplea la técnica conocida como Histograma de Multiresolución, consistente en emplear características calculadas a partir de histogramas obtenidos a distintas escalas de resolución de una misma imagen. De esta forma, se pretende tener en cuenta la información espacial de la imagen que con un único histograma no se tiene en cuenta. *Muhimmah et al. [15]* lo aplican en 2006 a la clasificación de la densidad de las mamografías, mediante un clasificador de Máquinas de Soporte Vectorial (SVM). Consiste en obtener una pirámide de imágenes reescaladas a la mitad de tamaño cada vez, y realizar varias operaciones sucesivas entre ellas para obtener un vector de características. Las operaciones son, en orden sucesivo: normalización de los histogramas, obtener los histogramas acumulativos, realizar la diferencia entre los histogramas acumulativos consecutivos y, por último, reducir cada una de las diferencias en un factor de $2^{3/4}$.



		MIAS Classification		
		F	G	D
Auto Class	F	97	19	5
	G	3	65	19
	D	6	20	87

Imagen 24. *Formación del vector de características y matriz de confusión*

Se obtiene una concordancia entre la máquina y el experto humano del 77,6 % usando la base de datos MIAS (con 3 clases). Aquí se probó que la información espacial proporcionada por los histogramas de multiresolución puede dotar de una capacidad más discriminadora al análisis de histograma en las mamografías.

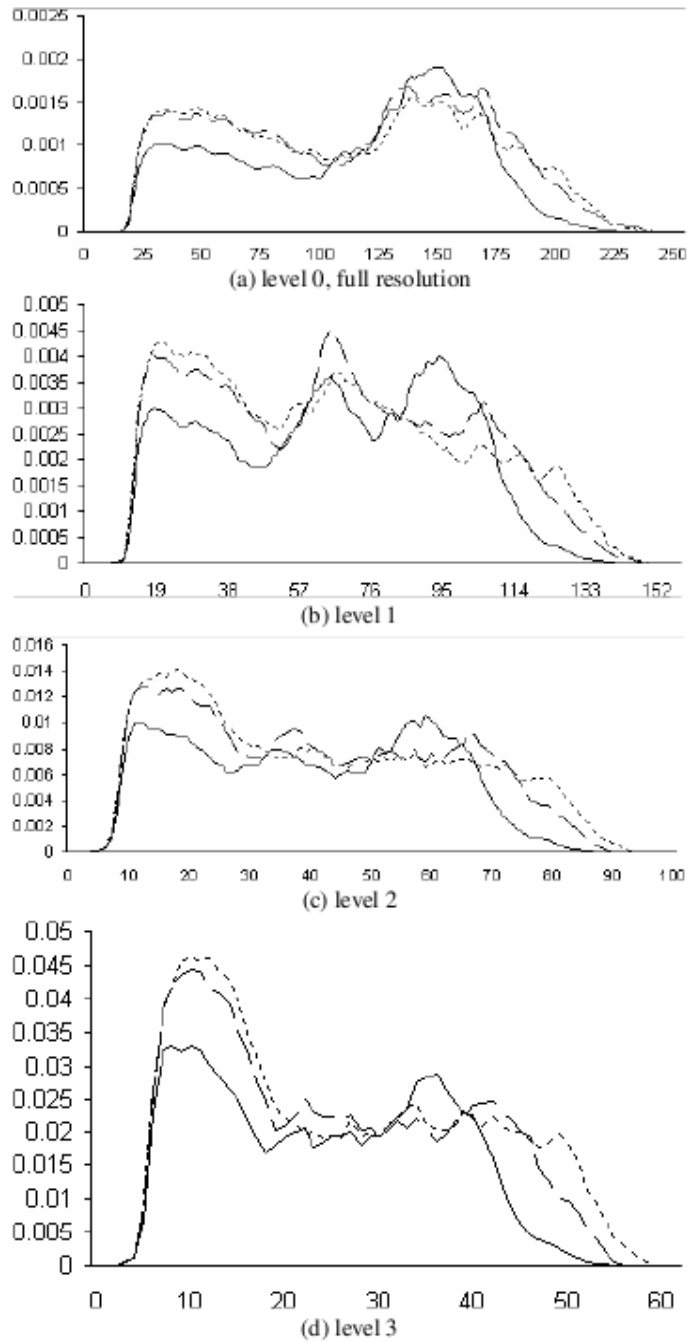


Imagen 25. Media de los histogramas de las 3 clases para los distintos niveles de reducción.

Bovis et al. evalúan cuatro aproximaciones distintas para extraer características de textura: matrices de niveles de gris utilizando distintas propiedades (momentos angulares, correlaciones, media, varianza, etc.), transformada de Fourier sobre la frecuencia de la mamografía, convolución con máscaras de texturas de Law y Transformación Discreta de Wavelets (DWT). Luego reducen el conjunto de características resultante mediante PCA. Y en la etapa de clasificación utilizan una combinación de clasificadores. Se evalúan 377 mamografías obtenidas de la base de datos DDSM. Los autores consiguen una tasa de aciertos del 71,4 % cuando emplean cuatro niveles de densidad, y del 96,9 % cuando se definen únicamente dos niveles de densidad.

(a)		(b)	
Combination Method	Recognition Rate %	Combination Method	Recognition Rate %
Best single result	58.3	Best single result	77.3
Min rule	40.3	Min rule	90.4
Max rule	58.5	Max rule	90.4
Product rule	71.4	Product rule	96.9
Majority vote	62.7	Majority vote	89.1
Median rule	61.2	Median rule	93.1
Sum rule	69.5	Sum rule	96.7

Imagen 26. Resultados para cuatro (a) y dos (b) niveles de gris, dependiendo del método de combinación de clasificadores empleado.

Los excelentes resultados con 2 clases no sirvieron para ocultar los problemas que tenía la metodología a la hora de discriminar en un problema con más clases.

Saha *et al.* (2001) [12] presentan un método basado en conectividad difusa sobre diversos conjuntos de características. La conectividad difusa establece “enlaces” entre categorías de objetos, de manera que un objeto está más fuertemente conectado con una categoría cuando la longitud del enlace es menor.

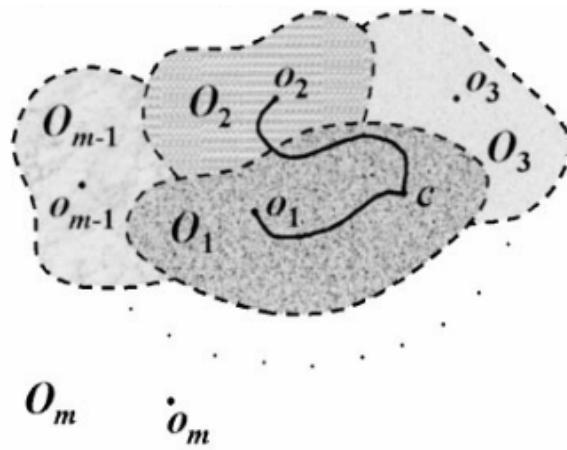


Imagen 27. Enlaces o1 y o2 del objeto c con las categorías O1 y O2

Demuestran que para establecer correctamente la densidad de la mama se requiere de técnicas más sofisticadas que el simple establecimiento de un umbral de gris.

En 2008, Oliver *et al.* [13, 14] presentan un nuevo método automático (perfeccionando uno que ellos mismos habían presentado) en el que dividen la mama automáticamente en tejido denso y no denso, mediante clustering (con *fuzzy C-means*). Con estos dos clusters, se obtienen características morfológicas y de texturas para estas dos clases: los cuatro primeros momentos del histograma para ambos clusters (como características morfológicas) y las matrices de coocurrencia (para las texturas). Los cuatro momentos del histograma utilizan la media, la desviación típica, asimetría y curtosis del histograma, mientras que las matrices de coocurrencia son matrices con las coocurrencias de pares de niveles de gris separados mediante un ángulo (que en el artículo cuenta con cuatro variaciones de 0°, 45°, 90° y 135°). Se utilizan dos clasificadores distintos, k-vecinos y árboles de decisión, para clasificarlas y por último se combinan ambos mediante una aproximación bayesiana de la forma que muestra la Imagen 28.

$$P(x \in B_c | A(x)) = \frac{P(A(x) | x \in B_c) P(B_c)}{\sum_{l=1..4} P(A(x) | x \in B_l) P(B_l)}$$

Imagen 28. Combinación Bayesiana de clasificadores

Los resultados con la base de datos MIAS (3 clases) obtiene un 86% de aciertos (utilizando las mamografías etiquetadas por tres expertos diferentes y anónimos) y con un subconjunto sin precisar de la DDSM (4 clases) consigue un 77%.

3.4. Aproximación propuesta

En 2008 presentamos un nuevo método para la clasificación de la densidad de mama [16]. En él, se dividen las mamografías de entrenamiento en subventanas de $N \times N$ píxels y se obtienen sus histogramas locales. Se aplica una reducción PCA a los subhistogramas y se clasifican en C clústers diferentes mediante k-means. Posteriormente, las mamografías se clasifican según el porcentaje de subventanas que pertenecen a cada clústers utilizando k-vecinos. Por ejemplo, al llegar una nueva mamografía al sistema clasificado mediante 3 clústers con k-means, la mama se subdividirá en ventanas y estas se designarán a cada uno de los 3 clústers, resultando por ejemplo que el 34% de las ventanas pertenecen al clúster 1, el 60% al clúster dos y el 6% restante al clúster 3. El vector de características sería de 3 componentes con los porcentajes de cada clúster. Al clasificarla con k-vecinos, el resultado sería el de la clase cuyos porcentajes sean más aproximados a los de la imagen.

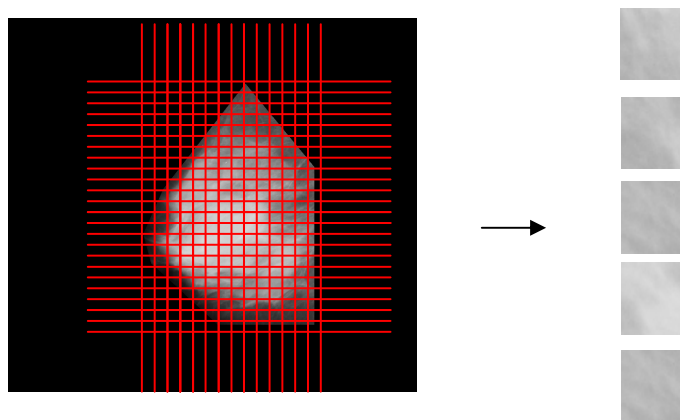


Imagen 29. Proceso de subdivisión en ventanas

Se obtuvo un resultado del 78% de concordancia con los expertos utilizando 6 clústers. Una comparativa con otros métodos utilizados comprobó que esta aproximación reducía el *Major Error* (confusiones entre las clases extremas, que son los errores más graves) respecto al resto (más información en el Capítulo 3.5).

	Fatty Glandular Dense			Author	Agreement (%)	Major errors (%)
Fatty	94	8	3	Our approach	77.96	1.24
Glandular	11	67	19	Muhimmah et al.	77.57	3.43
Dense	1	29	90	Oliver et al.	70.0	4.44
				Masek et al.	62.42	-

Imagen 30. Matriz de confusión y comparativa con otros métodos.

Por lo que se puede afirmar que actualmente nos encontramos con un error del 25-30% a la hora de clasificar la densidad de la mama. Sin embargo, es preciso añadir que la clasificación humana de las radiografías se realiza mediante expertos. Cada uno de ellos tiene su propio criterio y se pueden observar discrepancias entre ellos a la hora de clasificar un mismo caso. Si tenemos en cuenta que la clasificación automática se basa en la clasificación que ha realizado uno o varios expertos humanos, se puede afirmar que la tasa de error que se obtenga siempre va a depender del experto/expertos que hayan clasificado las muestras de entrenamiento, añadiendo así más variabilidad e incertidumbre a los resultados que se obtengan. Así, aunque existen varias bases de datos públicas con un único criterio de clasificación, una nueva clasificación realizada por otras personas podría hacer que los resultados empeoraran o mejoraran dependiendo de cómo se ajuste la nueva clasificación a los resultados de la clasificación automática.

4. Experimentación

En el presente trabajo se analizarán diferentes formas de obtener características de las mamografías para su posterior análisis. Se empezará con preprocesados básicos (como normalizaciones, ecualizaciones, etc.) y luego se implementarán algunas de las opciones descritas por los investigadores que han dado buenos resultados. Por último, se expondrá una metodología propia y se comparará con el resto. La base de datos MIAS viene con una clasificación de densidad en tres clases: *granular (G)*, *fatty (F)* y *Dense (D)*. En esta clasificación las mamografías más densas corresponden a la clase D y las menos densas a la clase F, estando la clase G en una situación intermedia entre las clases extremas. En esta distribución hay 104 mamografías G, 112 D y 106 de la clase F, todas a la misma resolución y digitalizadas con el mismo aparato. Por su parte, la base de datos DDSM dispone de unas 5000 imágenes clasificadas en cuatro clases 1,2,3,4. En este caso las imágenes tienen tamaños diferentes y han sido digitalizadas mediante diversos aparatos. Se pueden observar más detalles de las bases de datos públicas utilizadas en los apéndices sobre MIAS y DDSM.

Como puede verse en las imágenes mostradas a continuación, determinar el tipo de densidad de una mamografía no es sencillo, ya que aunque hay casos con los que utilizando simplemente el histograma puede deducirse que son diferentes, en otros muchos su histograma es similar...

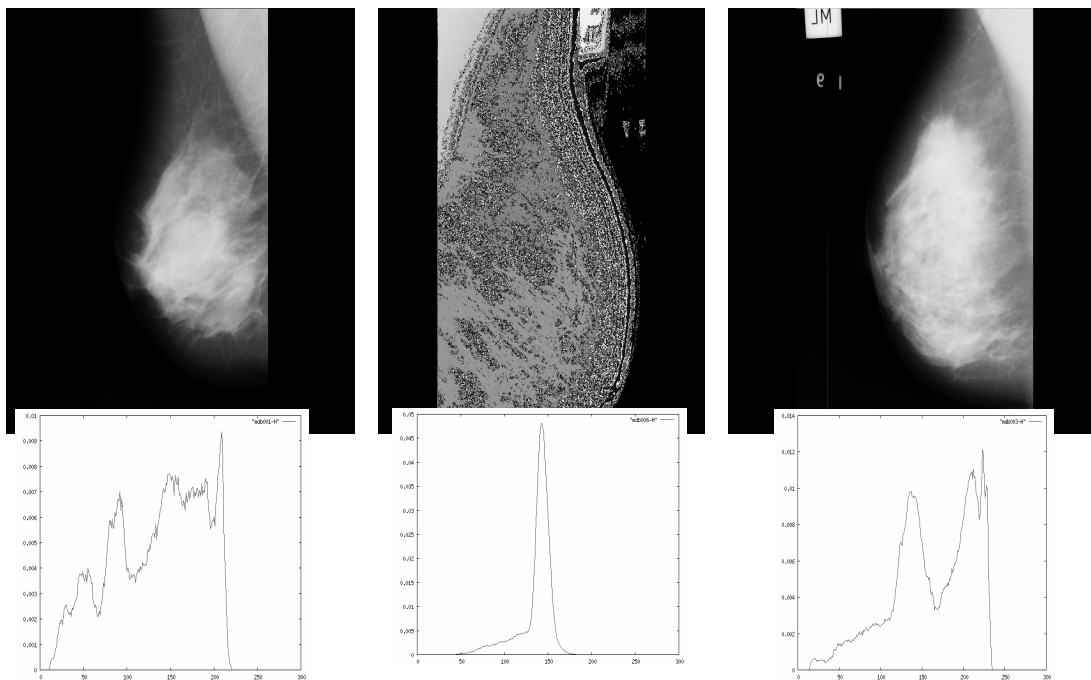


Imagen 31. Mamografías de clase G, F y D con sus histogramas diferentes.

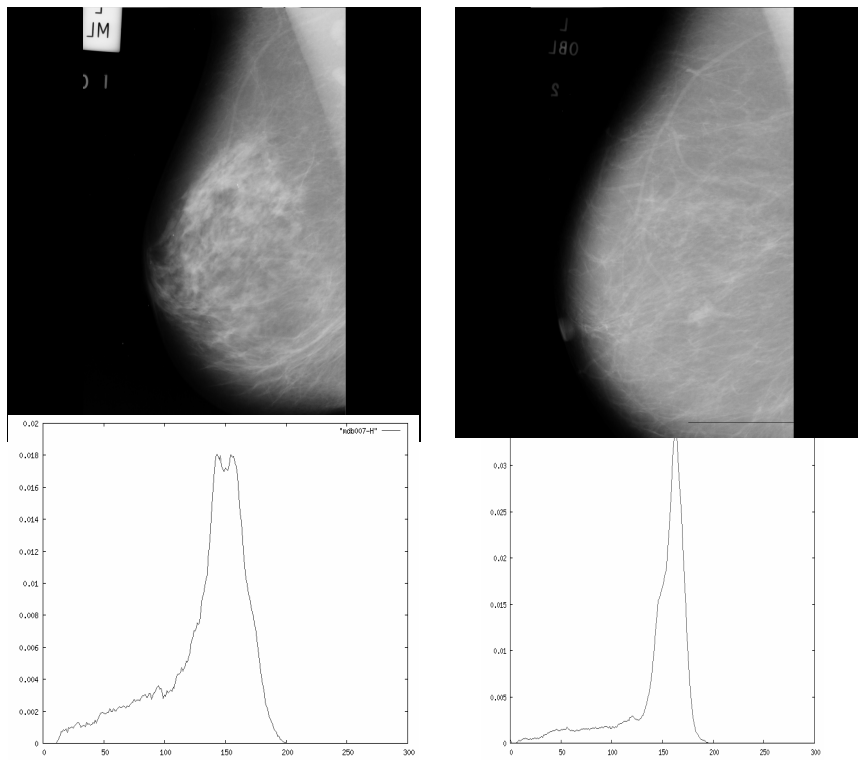


Imagen 32. Mamografía G y F con histogramas similares.

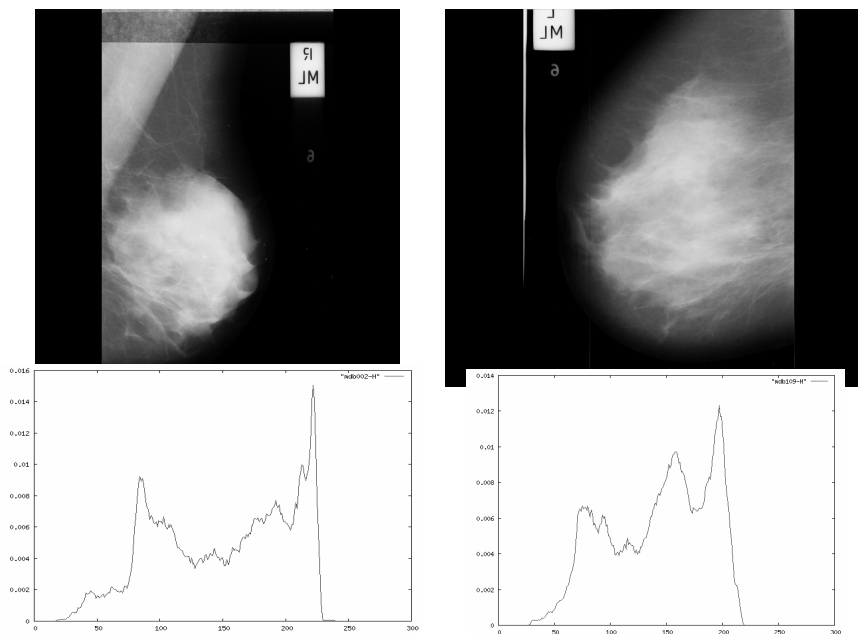


Imagen 33. Mamografías G y D con histogramas similares.

4.1. Preprocesado de la imagen

Aunque se disponía de una versión de la MIAS segmentada manualmente (sin fondo ni músculo pectoral), se ha optado por implementar un método de segmentación automática para posibles estudios posteriores o para segmentar mamografías originales. En una primera aproximación, se binariza la mamografía con un valor bajo de umbral (para no

eliminar ninguna información importante) y se la somete a un algoritmo de reconocimiento de objetos para quedarse con la mama (el objeto mayor).

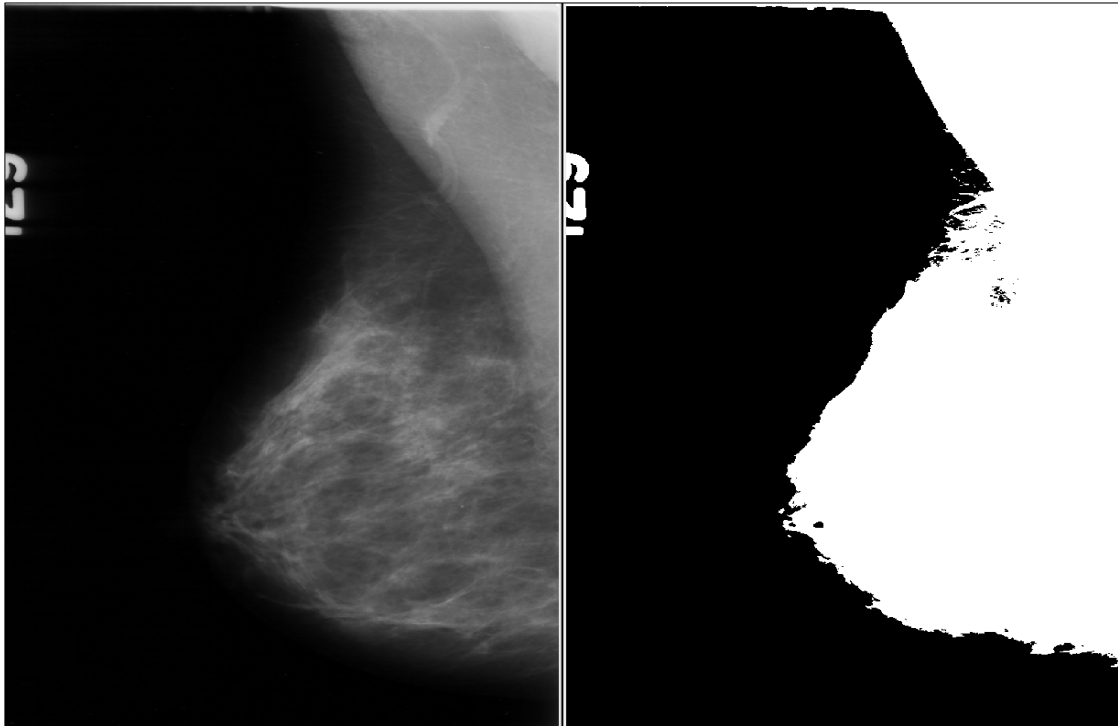


Imagen 34. Binarización de la imagen.

Sobre el objeto mayor se aplican las operaciones básicas de Apertura y Cierre para rellenar los posibles píxeles negros que se encuentren dentro del objeto (debidos a la binarización). Para eliminar los posibles errores en la mamografía (que pueden apreciarse en la Imagen 34 como una línea blanca clara en la parte superior), se hace un estudio de los niveles de gris presentes en los cuatro bordes de la imagen. Es en estas zonas donde se pueden presentar, y ante un nivel anormalmente elevado de gris se aplica una umbralización inversa sobre la zona (dejando esos valores excesivamente claros a color negro -fondo de la imagen-). Mediante este sencillo procedimiento se delimita la zona de la imagen que interesa estudiar y se pasa a separar el músculo pectoral de la misma.

Inicialmente, se realizaron varias pruebas con el rellenado de área. Se usa como punto de partida el píxel superior de la mama de la parte recta, y se va rellenando el área (que forma parte del músculo pectoral). El principal problema e inconveniente de esta técnica es determinar el umbral donde debe detenerse el crecimiento del área para que no se quede corta (dejando músculo pectoral en la imagen) ni se pase (eliminando parte de la mama en el proceso).

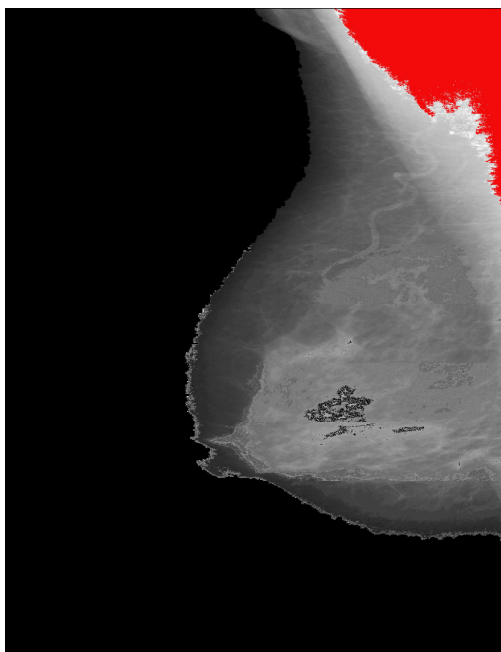


Imagen 35. Crecimiento de área que se ha detenido antes de eliminar completamente el músculo pectoral.

Tras varios intentos frustrados, se decidió implementar la aproximación de Hough para detectar y eliminar el músculo pectoral. Mediante la transformada de Hough se pueden encontrar patrones en las imágenes como rectas o circunferencias. Para n puntos de una imagen, se quieren encontrar los subconjuntos de puntos que caen en líneas rectas. La transformada de Hough puede acelerar el proceso que supondría hacer una búsqueda exhaustiva de todas las rectas que forman cada par de puntos y encontrar los puntos que forman parte de cada una.

Al considerar un punto (x_1, y_1) y la ecuación de la recta $y_1 = a \cdot x_1 + b$, se sabe que por el punto (x_1, y_1) pasan infinitas rectas que satisfacen la ecuación anterior para diferentes valores de a y b . Si se cambia la ecuación por $b = -x_1 \cdot a + y_1$ y se considera el plano ab (llamado espacio de parámetros), se tiene una única recta para el punto (x_1, y_1) . Otro punto (x_2, y_2) también tendrá su recta en el espacio de parámetros. Las rectas de (x_1, y_1) y (x_2, y_2) se cortarán en un punto (a', b') , donde a' será la pendiente y b' la ordenada al origen de la recta que contiene a ambos puntos (x_1, y_1) y (x_2, y_2) .

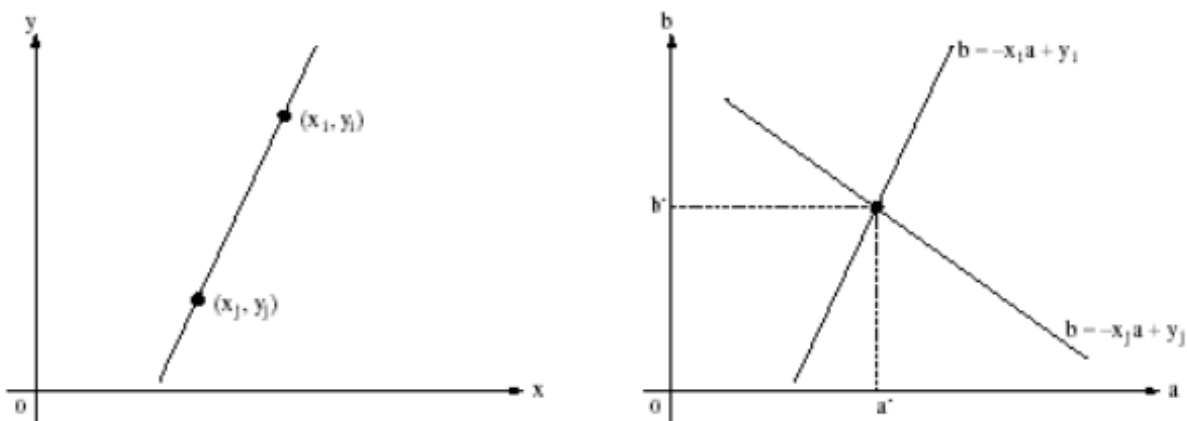


Imagen 36. Izquierda: Plano xy ; derecha: Plano ab (Espacio de parámetros).

Para la transformada de Hough, el espacio de parámetros se divide en celdas

“acumuladoras”, como se puede ver en la siguiente imagen. En ella a_{min} , a_{max} , b_{min} , b_{max} son los rangos esperados de la pendiente y la ordenada al origen. Inicialmente, todos los acumuladores se ponen a 0. Luego se calcula para cada punto (x,y) de la imagen el valor b mediante la ecuación anterior ($b=-x1*a+y1$). A un valor a_p le corresponde un valor b_q , y el acumulador $A(p,q)$ pasa a valer $A(p,q)+1$. Por lo tanto, un valor V en el acumulador $A(i,j)$ significa que V puntos del plano xy caen sobre la recta $y=a_i*x+b_j$.

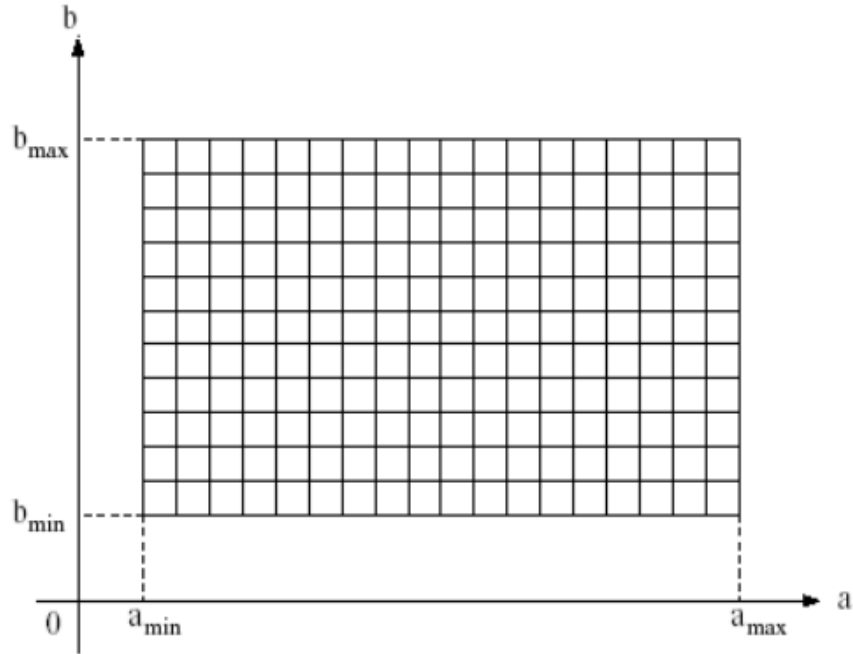


Imagen 37. Celdas acumuladoras en el espacio de parámetros.

Un problema que puede aparecer al usar la ecuación de la recta $y=a*x+b$ es que tanto la pendiente como la ordenada al origen pueden llegar a tomar un valor infinito, a medida que la línea se vaya haciendo vertical. Para evitarlo, en la práctica se usa la representación normal de la recta: $x*\text{Cos}(\theta)+y*\text{Sen}(\theta)=p$.

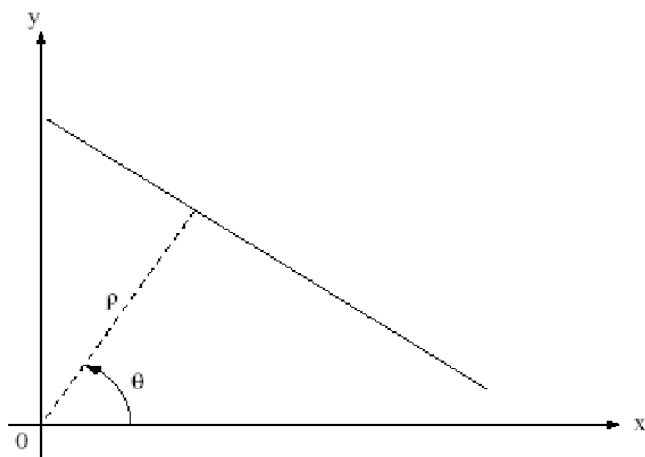


Imagen 38. Representación normal de la recta.

El uso de esta representación para construir la tabla es análogo al anterior, donde a cada punto del plano xy le corresponde una sinusoidal en el plano $0p$ (en lugar de una recta). Como antes, V puntos colineantes a la recta $x*\text{Cos}(\theta_i)+y*\text{Sen}(\theta_i)=p_i$ darán lugar a V sinusoides que se cortan en el punto $(p_i,0_i)$ del espacio de parámetros. Incrementando θ y calculando p se obtendrán V entradas en el acumulador $A(i,j)$.

El espacio de parámetros se verá modificado ligeramente con la representación normal. El rango para θ es de $[-90^\circ, 90^\circ]$ respecto al eje de abscisas y ρ puede tener valores negativos.

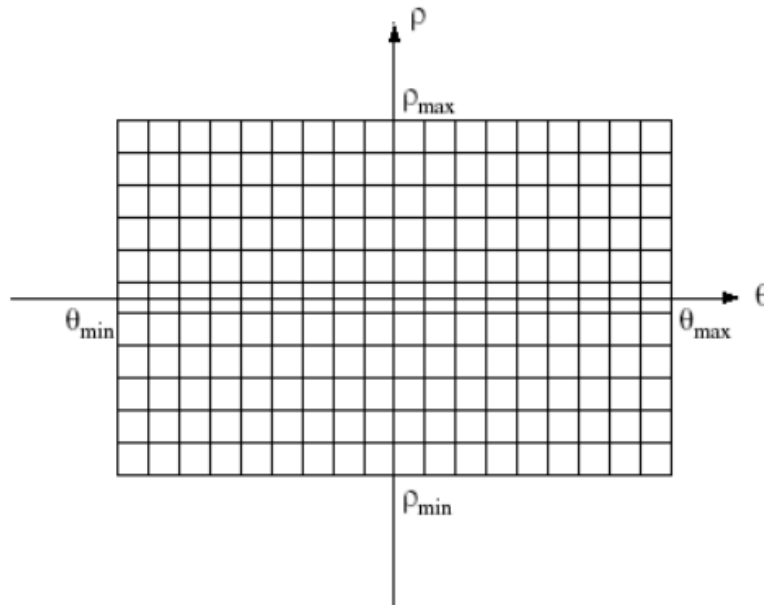


Imagen 39. Nuevo espacio de parámetros.

En la siguiente imagen se presenta un ejemplo sencillo de la transformada de Hough. Se tiene una imagen con 4 puntos. Al pasarla a la transformada de Hough se obtienen 4 sinusoidales en el espacio $\rho\theta$, que se cortan en seis puntos distintos (en la imagen se cortan en ocho puntos, pero hay que observar que los puntos con $\theta=90^\circ$ son iguales que los de $\theta=-90^\circ$) correspondientes a las seis rectas posibles que pasan por los cuatro puntos de la imagen (los cuatro lados del cuadrado y las dos diagonales).

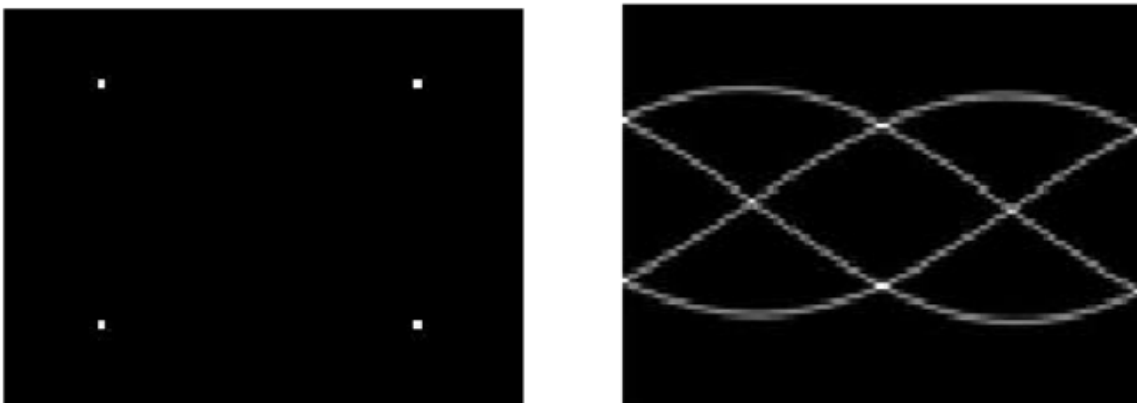


Imagen 40. Imagen con cuatro puntos y su correspondiente Transformada de Hough.

4.2. Extracción de características y experimentación

La experimentación ha consistido en extraer características de las imágenes de la base de datos MIAS de diferentes maneras para comprobar cómo afecta a la clasificación de la densidad de la mama. La MIAS consta de 322 mamografías clasificadas en tres clases (G, D y F), y contiene mamografías de la mama derecha e izquierda de los pacientes (para más información, ver el apéndice de las Bases de Datos utilizadas).

Se usa un clasificador de k-vecinos utilizando una estrategia denominada *leaving one group out* para aprovechar al máximo las imágenes de la Base de Datos. Con esto, cada una de las imágenes se clasifica utilizando las 320 restantes como *training* (de las 321 se quita la otra que corresponde al mismo paciente), aumentando lo posible el n° de muestras de entrenamiento mientras que se evita entrenar con un posible duplicado (la perteneciente al mismo paciente) de la que se está clasificando. En algunos experimentos se ha aplicado también *Principal Component Analysis* (PCA) al conjunto de características para ver el impacto que produce sobre el resultado.

El PCA es un procedimiento mediante el que se transforma un espacio de datos en otro equivalente. Básicamente, usa transformaciones lineales ortogonales para pasar un sistema de coordenadas a otro. Mediante el PCA los vectores de características pueden ser reducidos a menos componentes y hacer más manejable la información sin perderla.

4.3. Resultados con histograma global normalizado

Como primera aproximación, se utilizó únicamente el histograma global normalizado de las mamografías como características para la clasificación. El histograma de una imagen es el número de píxeles que hay en cada uno de los niveles de gris (de 0 a 255, en este caso). El histograma normalizado se consigue haciendo que la suma de todos los valores sea 1 (dividiendo cada componente del histograma por el número total de píxeles de la imagen).

Se utilizó k-vecinos variando k por todo el rango de vecinos posibles (de 1 a 322), y se obtuvo el mejor resultado con 24 vecinos y un error total del 31.99 %.

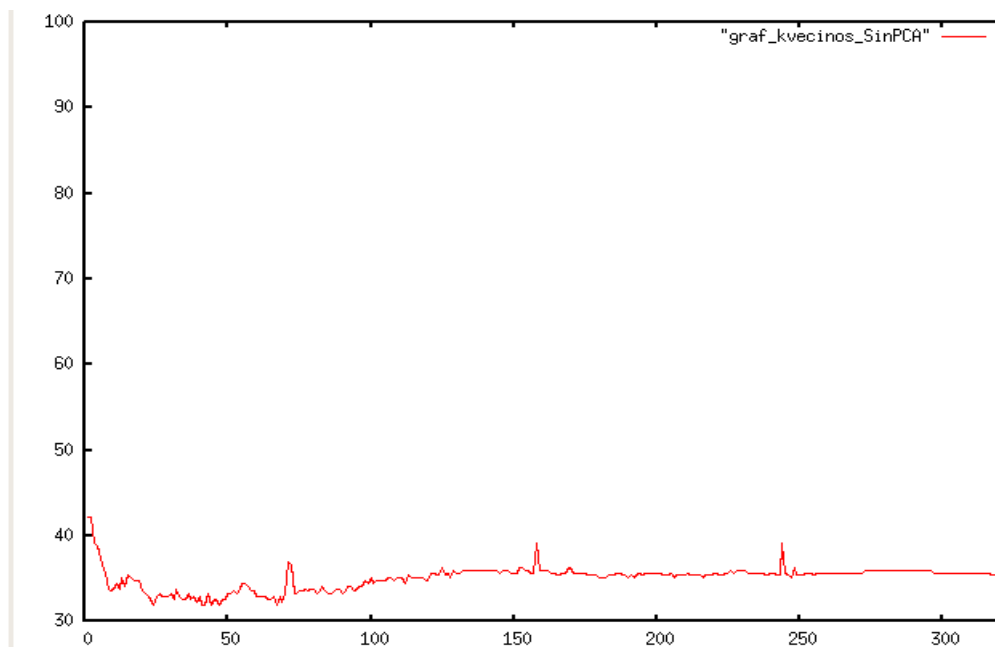


Imagen 41. N° de vecinos (x) y porcentaje de error (y).

A continuación se empleó una reducción del histograma para ver si se conseguía mejorar el resultado obtenido. Un histograma se reduce redimensionando los niveles de gris y distribuyendo sus píxeles entre los nuevos niveles (por ejemplo, pasar de 256 niveles de gris a 20, 30 o 100).

Se usaron las mismas características que antes (los histogramas) pero reduciéndolos el rango de 2 a 50, clasificando con k-vecinos variando k entre los vecinos posibles.

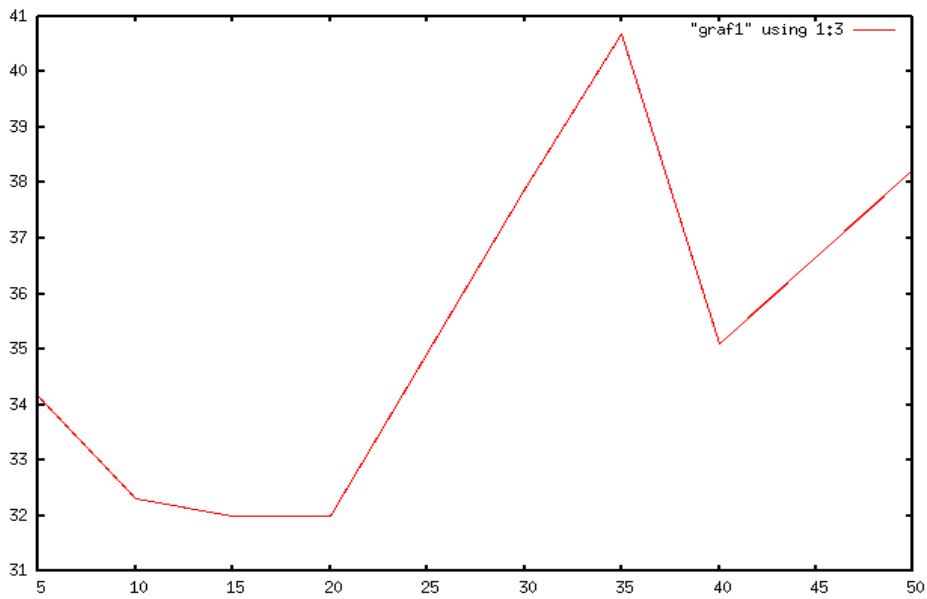


Imagen 42. Reducción (x) y porcentaje de error (y) para $k=20$ (nº de vecinos)

Se observó que los mejores resultados se obtenían reduciendo el histograma entre las 10 y 25 dimensiones, y se procedió a una experimentación más exhaustiva dentro de ese rango.

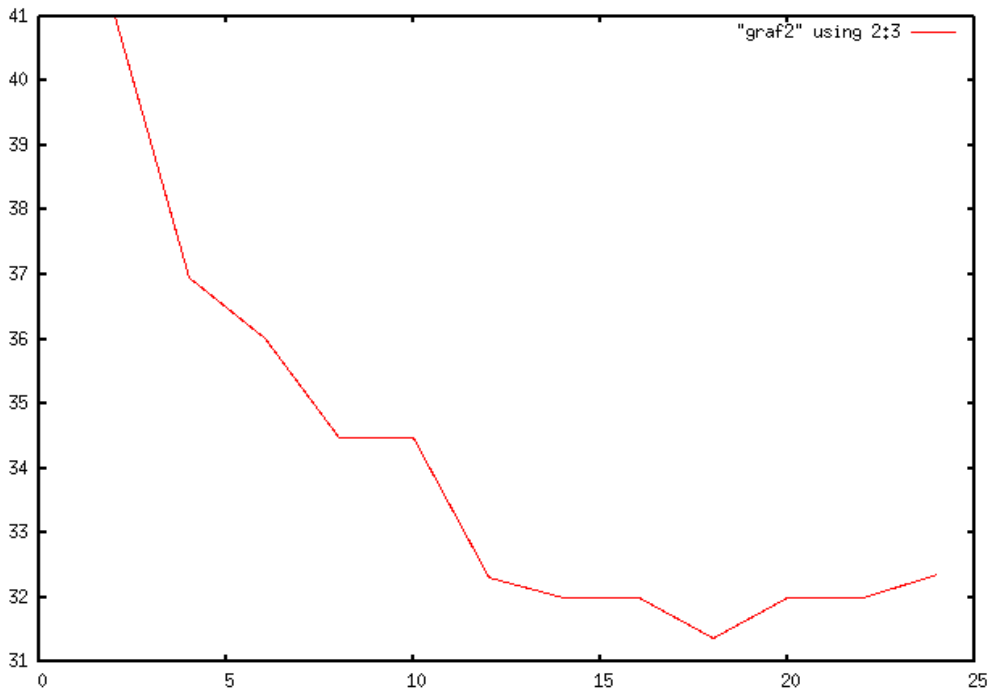


Imagen 43. Nº de vecinos (x) y porcentaje de error (y) para Suavizado 17

Las pruebas realizadas muestran que el mejor resultado se obtuvo con una reducción 17 y utilizando 18 vecinos, con un error del 31.37 %.

	F	G	D
F	78	14	5
G	24	71	35
D	4	19	72

TOT=322 OK=221 ERR=101; OK=68.633540 Err=31.366460

Imagen 44. Matriz de confusión para el Suavizado 17 y 18 vecinos.

Aunque la mejora es casi inapreciable (31.99 frente a 31.37), sí es recomendable trabajar con histogramas reducidos ya que es más eficiente trabajar con un histograma reducido (en este caso de 17 dimensiones) que con el histograma completo de 256 niveles de gris.

Luego se repitieron los experimentos, pero aplicando PCA sobre el conjunto de características para comprobar las diferencias entre los resultados que se obtenían. Usando el histograma global de las mamografías normalizadas, se probó kvecinos con el rango de vecinos posibles (de 1 a 322) y redimensionado pca en todo el rango posible (de 1 a 256 componentes).

Se obtuvo el mejor resultado en varias ocasiones: 9 componentes principales y 67 vecinos, 13 componentes principales y 27 vecinos, 51 componentes principales y 28 vecinos, 61 componentes principales y 67 vecinos, 62 componentes principales y 67 vecinos o 63 componentes principales y 28 vecinos; todos con un error total del 31.37%.

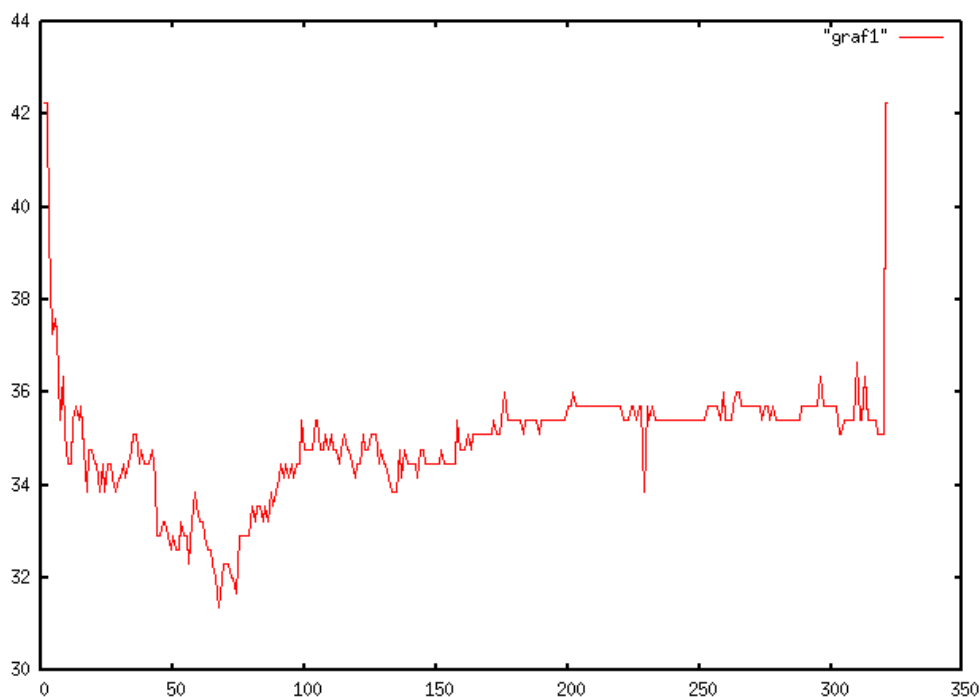


Imagen 45. N° de vecinos (x) y porcentaje de error (y) para un redimensionado PCA a 9 dimensiones

Al igual que antes, se repitió el experimento pero añadiendo una reducción al histograma (antes de aplicar PCA) con el mismo intervalo (2 a 50). Aquí se pudo comprobar que la reducción no afecta al resultado con PCA, ya que se obtuvo el mismo resultado que sin usarlo en dos configuraciones diferentes: redimensionado 17, 12 componentes principales y 18 vecinos o redimensionado 17, 14 componentes principales y 18 vecinos, con el mismo resultado del 31.37% de error.

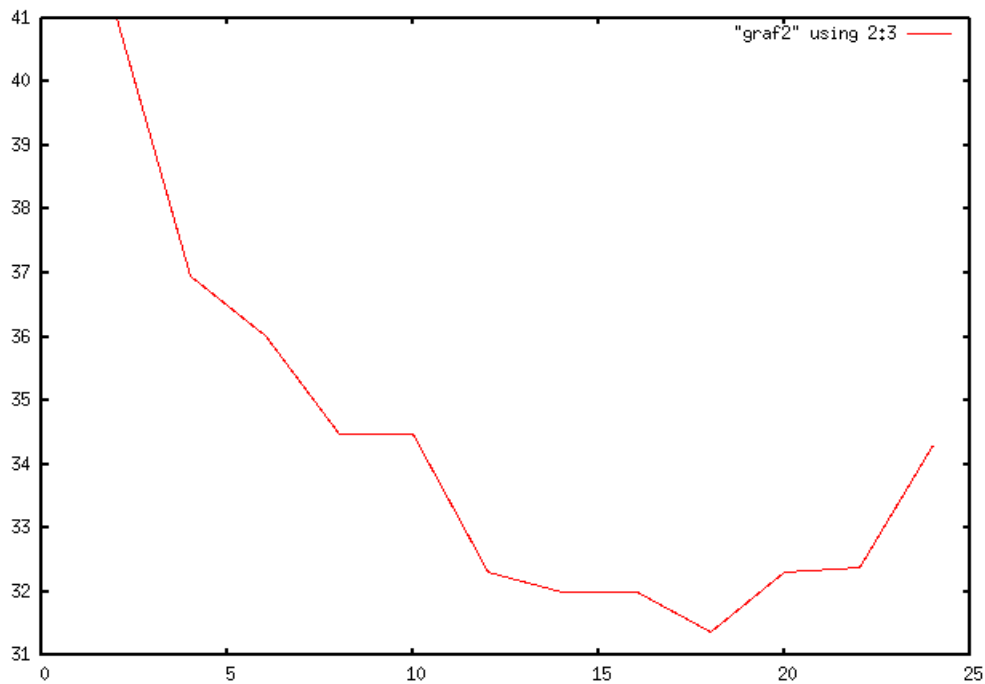


Imagen 46. N° de vecinos (x) y porcentaje de error (y) con redimensionado 17 y 12 componentes principales

Se obtiene un 31.37% de error usando únicamente el histograma global de las mamografías segmentadas. El uso de PCA no consigue mejorar significativamente el resultado en este caso, ya que se obtiene el mismo porcentaje de error que al realizar una reducción del histograma sobre los datos originales. También se probó una cuantización previa de los niveles de gris de la imagen en lugar de la reducción del histograma, pero no se consiguió mejorar.

Por lo tanto, podemos establecer un error máximo del 32%. Cualquier técnica más compleja que el cálculo de un histograma y su reducción será inválida si se obtiene un porcentaje de error mayor que el obtenido en estos casos.

4.4. Resultados al aplicar diversas técnicas sobre el histograma global

MULTIRESOLUTION HISTOGRAM:

Técnica usada en el artículo de *Muhimmah et al.* y explicada en el de *Hadjidemetriou et al.* [15]. Esta reducción presentó algunos problemas ya que el número de componentes correspondiente es de 582 (reducciones 256+161+101+64), mientras que en el artículo indicado se usa un total de 580 (reducciones 256+160+96+68) componentes, así que se usaron las dos reducciones posibles para contrastar resultados.

El artículo muestra unos resultados del 22.43% de error usando un clasificador DAG-SVM. Aquí se comprobaron sus prestaciones con k-vecinos y la base de datos MIAS.

Al usar la técnica de Multiresolution Histogram con k-vecinos (recorriendo el rango posible de 1 a 322), se obtuvieron resultados bastante pobres.

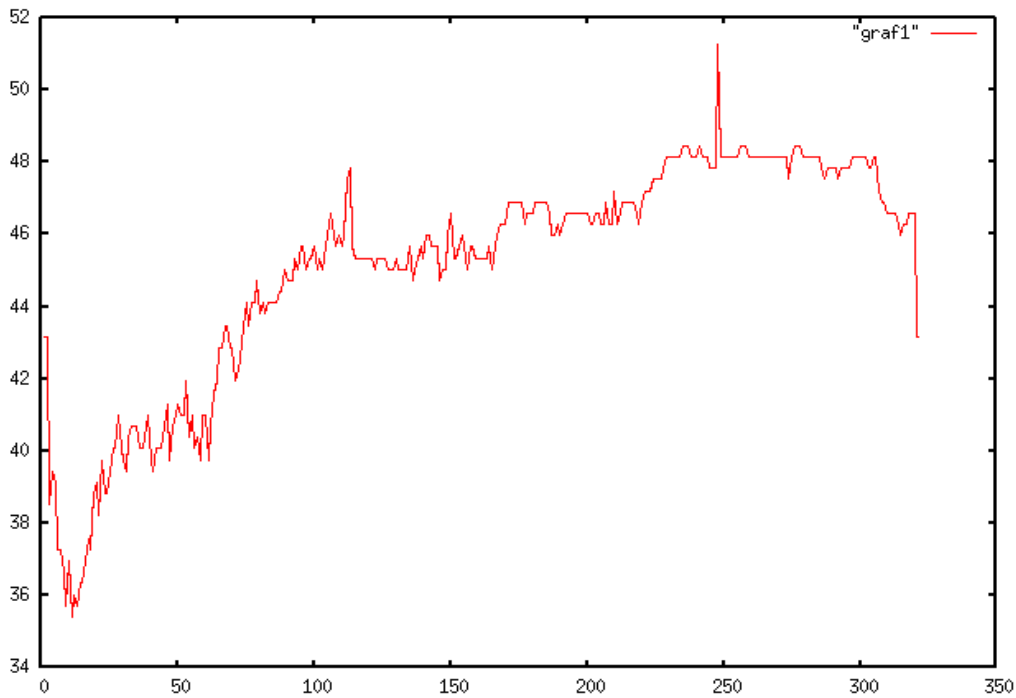


Imagen 47. N° de vecinos (x) y porcentaje de error (y) obtenido con 582 características.

El mejor resultado con 582 características fue del 35.4% de error con 11 vecinos.

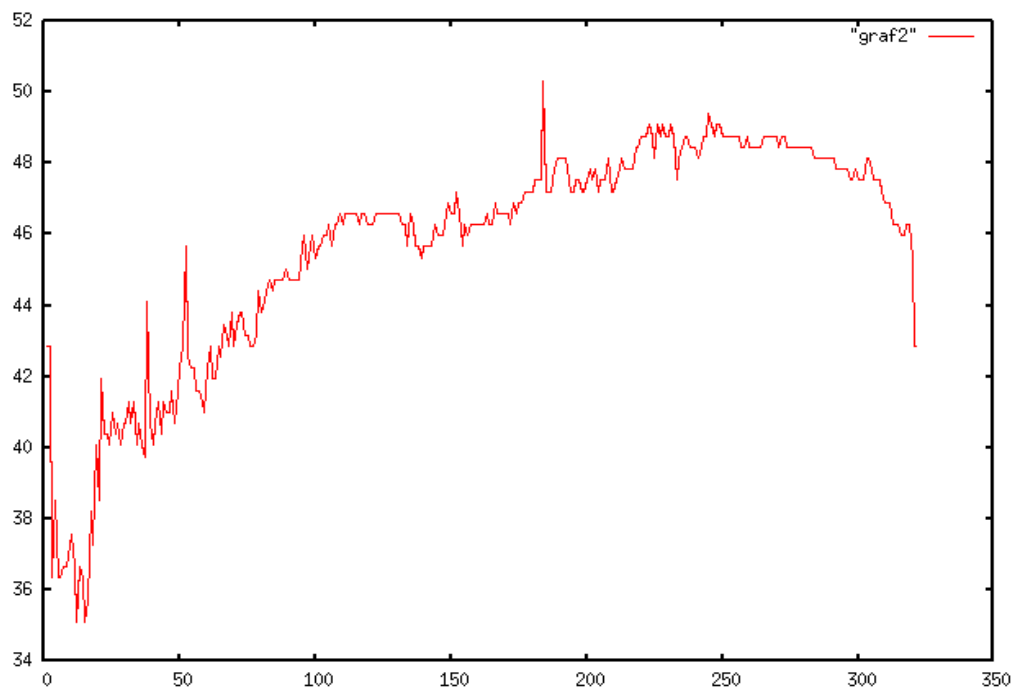


Imagen 48. N° de vecinos (x) y porcentaje de error (y) obtenido con 580 características.

Con 580 características, se obtuvo como mejor resultado un valor del 35.09% de error con 12 vecinos.

De nuevo, se repitió el experimento aplicando PCA sobre las características anteriores, mejorando el resultado pero sin producirse avances importantes respecto a no aplicar la técnica del artículo. Reescalando las dimensiones entre 1 y 582 (o 580, según el conjunto de características usado), se obtuvo el mejor resultado en dos configuraciones diferentes:

con 21 vecinos y 4 dimensiones, o 32 vecinos y 8 dimensiones con un error del 32.3% (para el conjunto de 582 características).

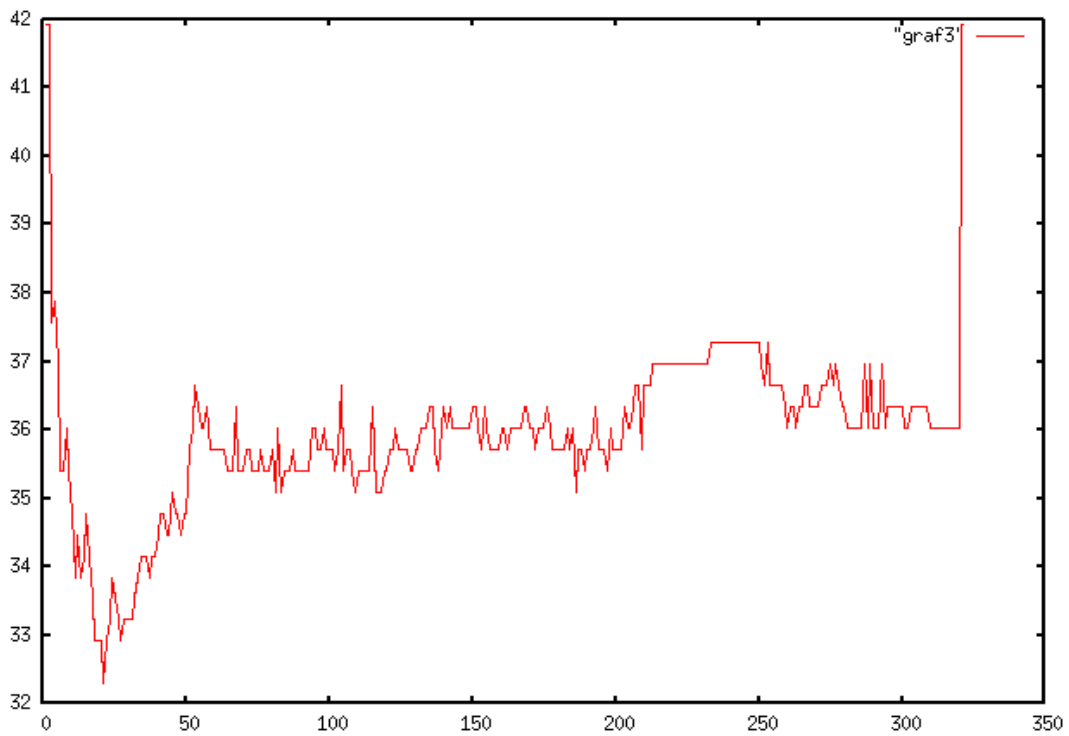


Imagen 49. N° de vecinos (x) y porcentaje de error (y) para 4 dimensiones y 582 características.

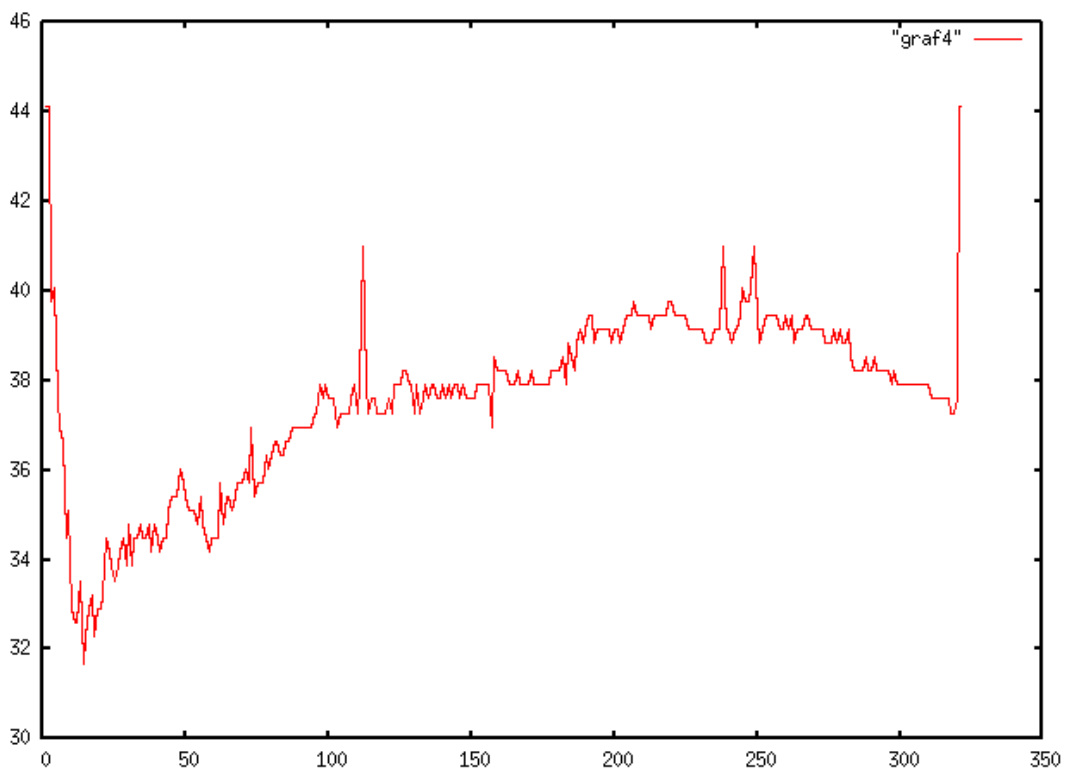


Imagen 50. N° de vecinos (x) y porcentaje de error (y) para 5 dimensiones y 580 características.

Con el conjunto de 580 características se obtuvo un 31.68% de error con 5 dimensiones y 14 vecinos.

Por último, se probó a utilizar las imágenes segmentadas manualmente (y no el resultado de la segmentación automática) para el experimento, usando 580 características por ser la configuración que “mejores” resultados presentaba. Aunque debería dar los mismos resultados que al usar la automática, se consiguieron unos ligeramente mejores.

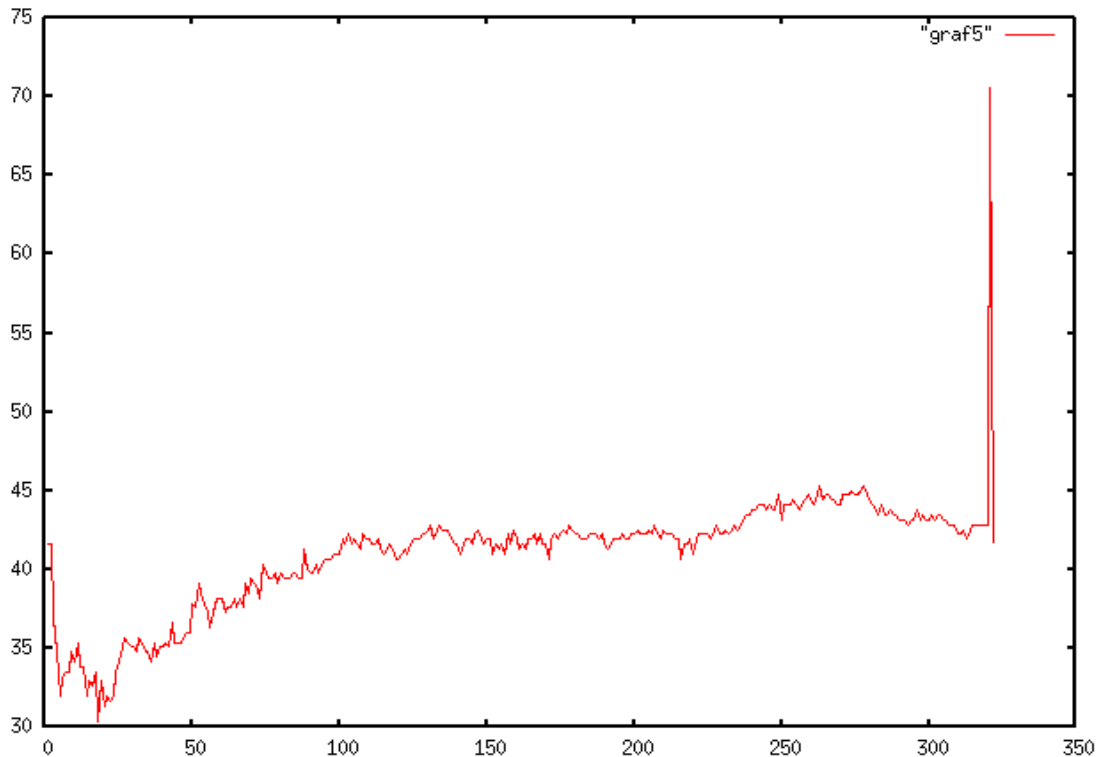


Imagen 51. *Nº de vecinos (x) y porcentaje de error (y) para 30 dimensiones con 580 características, usando directamente imágenes segmentadas.*

El mejor resultado obtenido en este caso fue del **30.43% de error** con 30 dimensiones, 18 vecinos, 580 características y usando las imágenes segmentadas manualmente.

HISTOGRAMA POR REGIONES:

La técnica viene del artículo de *Karssemeijer [11]*, que obtiene un porcentaje de acierto del 65% (35% de error) con una base de datos desconocida formada por un conjunto de test de 615 casos.

Primero se procedió a realizar el experimento (sin aplicar PCA) usando k-vecinos en todo el rango posible (de 1 a 322). Los resultados, de nuevo, no mejoraron el *baseline* obtenido con el histograma global (capítulo anterior). Se obtuvo como mejor resultado un porcentaje del 32.92% de error con 13 o 21 vecinos.

Como de costumbre, se repitió el experimento aplicando PCA y se obtuvo el mejor resultado al utilizar 10 dimensiones y 19 vecinos, 20 dimensiones y 31 vecinos o 100 dimensiones y 31 vecinos, un porcentaje de error del 30.12%

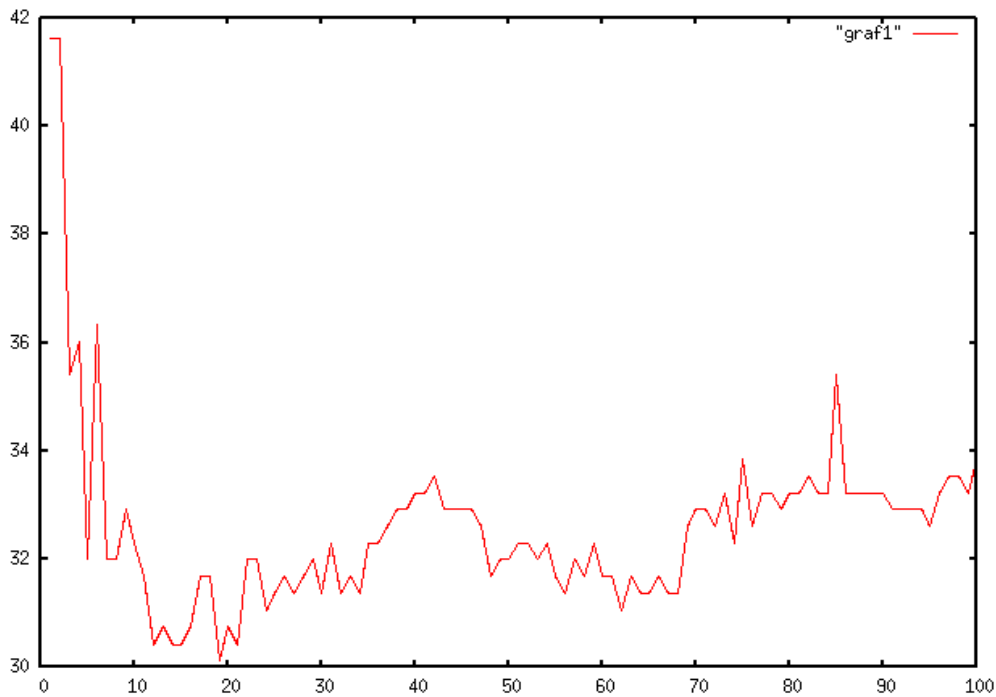


Imagen 52. N° de vecinos (x) y porcentaje de error (y) para 10 dimensiones.

De forma análoga a lo realizado en la técnica anterior, se procedió a usar directamente las imágenes segmentadas manualmente (sin músculo pectoral ni ruido de fondo) para calcular la segmentación de la mama, mejorando ligeramente el resultado obtenido hasta el momento. El mejor resultado con este experimento se obtuvo con dimensión 6 y 40 o 41 vecinos y con dimensión 17 y 38 vecinos: un 29.5 % de error.

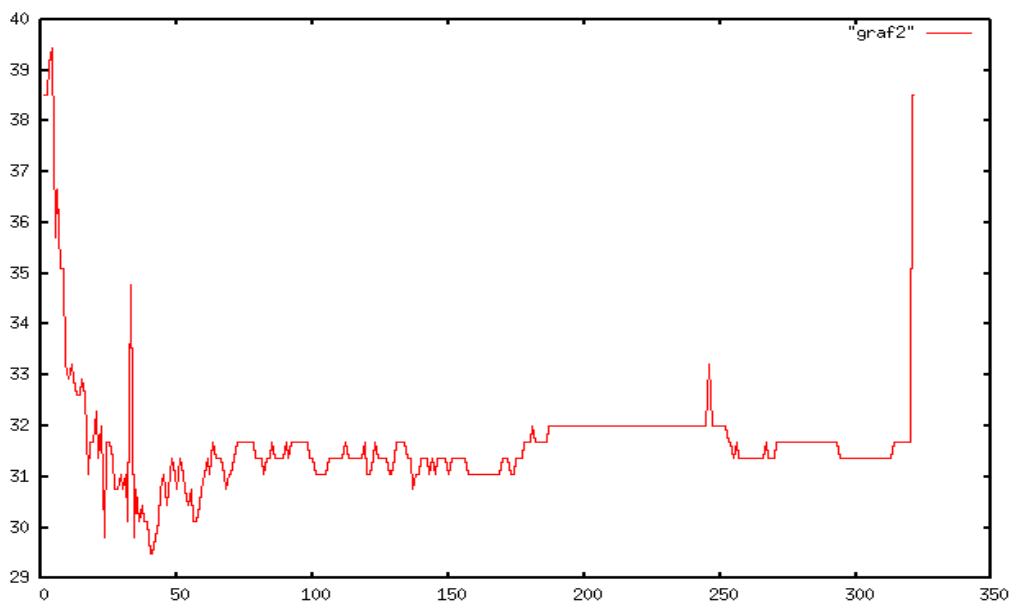


Imagen 53. N° vecinos (x) y porcentaje de error (y) para 6 dimensiones y usando la segmentación manual.

MULTIRESOLUTION HISTOGRAM E HISTOGRAMA POR REGIONES:

Se probó a combinar las dos técnicas anteriores para ver si se mejoraban los resultados, de modo que se realizaba el reescalado de las mamografías y se procedía a segmentar las

mismas en cinco partes, quedándose con las tres partes centrales para realizar el histograma y las operaciones descritas en el *Multiresolution Histogram*.

Sin embargo, los resultados fueron bastante peores. Variando los vecinos entre 1 y 322, y redimensionando el sistema de 100 a 1000 características se obtuvieron resultados pobres, no llegándose ni a igualar lo obtenido individualmente por cada experimento.

Se consiguió como mejor resultado un 36.65% de error con 61 vecinos y 200 dimensiones...

SELECCIÓN DE CARACTERÍSTICAS:

Por último, se usaron los ficheros de características con los que se habían obtenido mejores resultados y se implementó un algoritmo de selección de características basado en el algoritmo *Fast-forward* para ver cuáles ofrecían mejores resultados, primero usando las características de un único experimento e ir combinando características de entre todas las posibles (no sólo de entre un mismo experimento). De este modo, se elige una única característica de entre todas que ofrezca el mejor resultado de clasificación. Sucesivamente, se le va añadiendo la siguiente característica que combinada con las ya seleccionadas ofrezca el mejor resultado de clasificación posible.

·Con Multihistograma:

Se obtuvo el mejor resultado con 36 características seleccionadas (563 59 435 241 49 334 56 123 265 576 234 174 416 574 231 248 575 406 232 439 391 233 53 235 403 218 219 333 171 573 381 191 390 393 212 230) del multihistograma, con un error del 23.91%

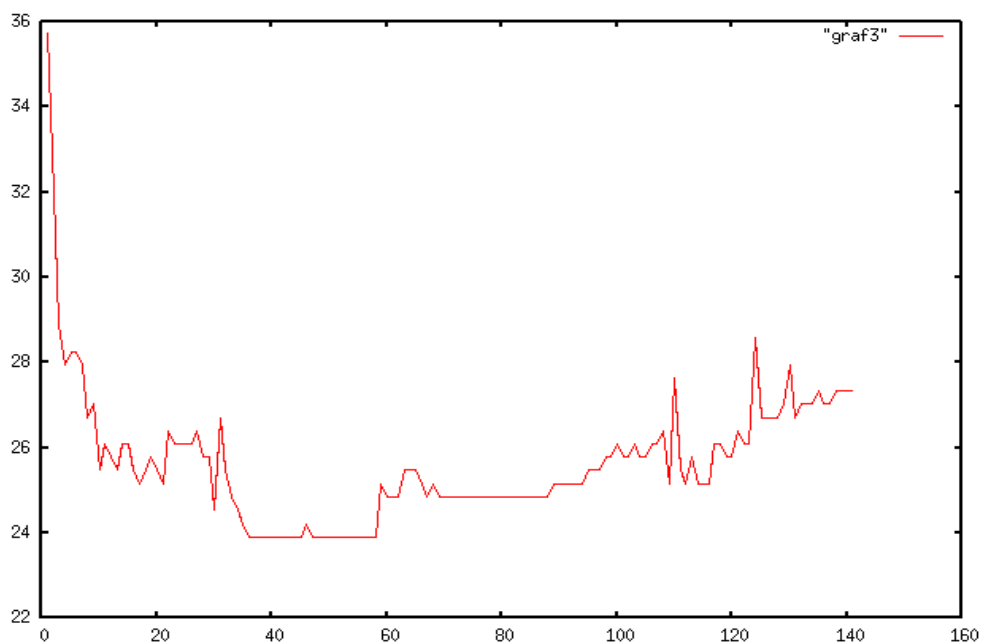


Imagen 54. Porcentaje de error (y) y nº de características (x) seleccionadas para el Multihistograma.

·Con Histogramas por regiones:

El mejor resultado se obtuvo al combinar 25 características (453 435 370 479 437 399 427 214 373 439 80 475 364 1 74 336 66 225 73 596 2 481 236 226 3) de Histograma por Regiones, con un porcentaje de error del 23.91 % (Igual que en el caso de Multihistogramas).

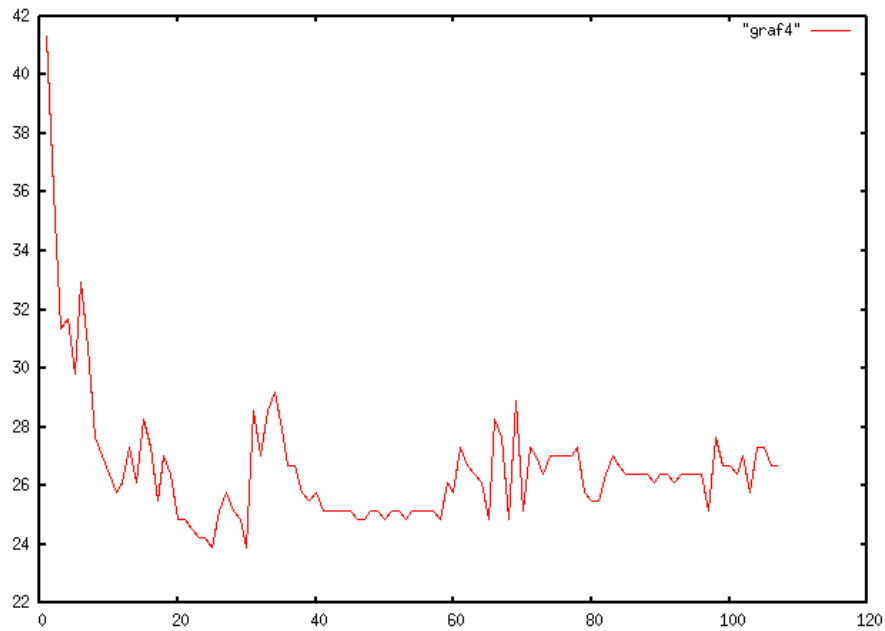


Imagen 55. Porcentaje de error (y) y nº de características (x) seleccionadas para Histograma por regiones.

·Con texturas:

Se utilizó un fichero de características obtenido mediante el uso de texturas (experimento no realizado por el autor) para observar las texturas más significativas, y se obtuvo el mejor resultado con 16 características seleccionadas: 126 29 114 1 91 128 37 134 45 133 33 39 40 138 77 48, un porcentaje de error del 30.75%.

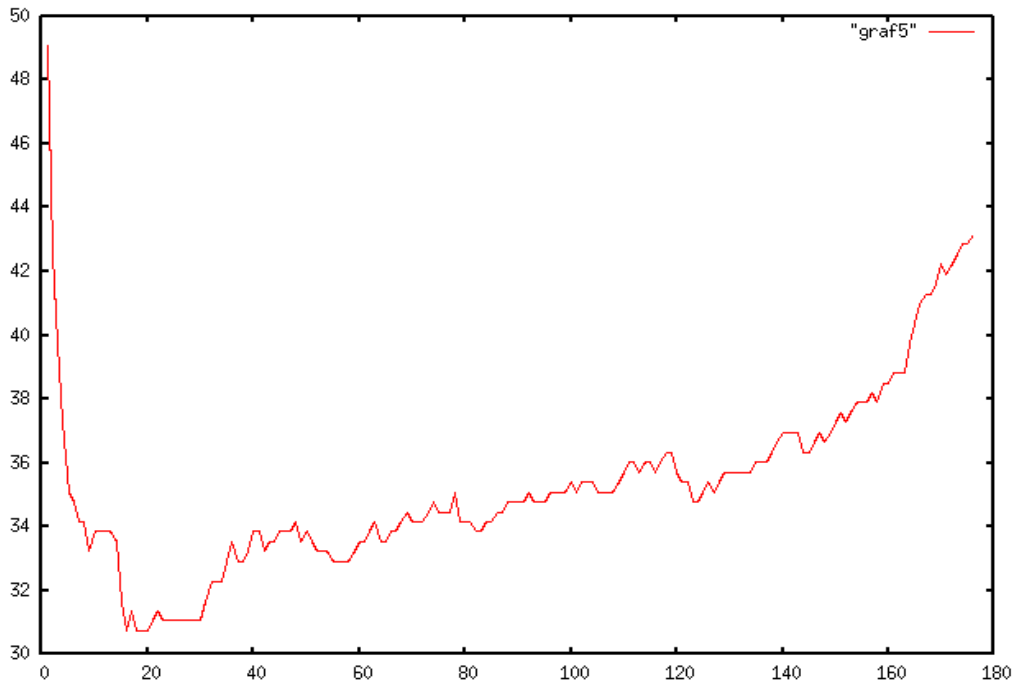


Imagen 56. Porcentaje de error y nº de características seleccionadas para texturas.

·Con histograma normalizado:

Se utilizó un fichero de características con el histograma global normalizado, obteniendo el mejor resultado con 78 características (198 136 307 402 229 160 493 123 361 155 227 247 248 249 250 251 252 253 401 234 398 244 235 236 239 390 327 230 456 403 399

241 242 243 245 246 254 255 256 407 409 410 411 412 413 414 415 416 417 233 237
 238 228 397 226 240 225 305 99 404 408 395 1 146 224 192 577 517 308 344 79 202
 231 406 322 303 578 86). En esta ocasión el porcentaje de error fue del 28.26 %.

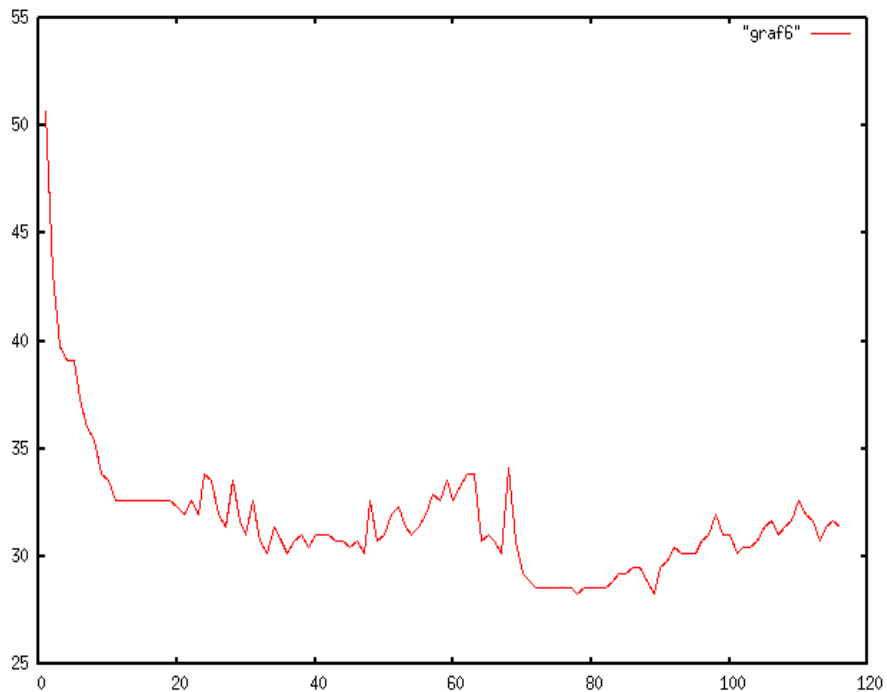


Imagen 57. Porcentaje de error (y) y n° de características (x) seleccionadas para el histograma normalizado.

· Combinación de los ficheros:

Por último, se usaron las mejores características de los experimentos anteriores en un único fichero, con la siguiente distribución: Texturas (características 1-16), MultiHistograma (características 17-52), K-means (características 53-54, obtenidas por otro investigador e incluidas de forma arbitraria en el estudio), Histograma por regiones (características 55-79). Se obtuvo el mejor resultado con 36 características (48 20 46 34 17 38 19 21 36 52 32 23 45 50 29 35 51 44 30 47 41 31 18 33 43 26 27 37 22 49 39 24 40 42 25 28) y un porcentaje de error del 23.91 %.

Las características seleccionadas pertenecen a las más significativas del MultiHistograma, sin que el resto consigan proporcionar mejora alguna...

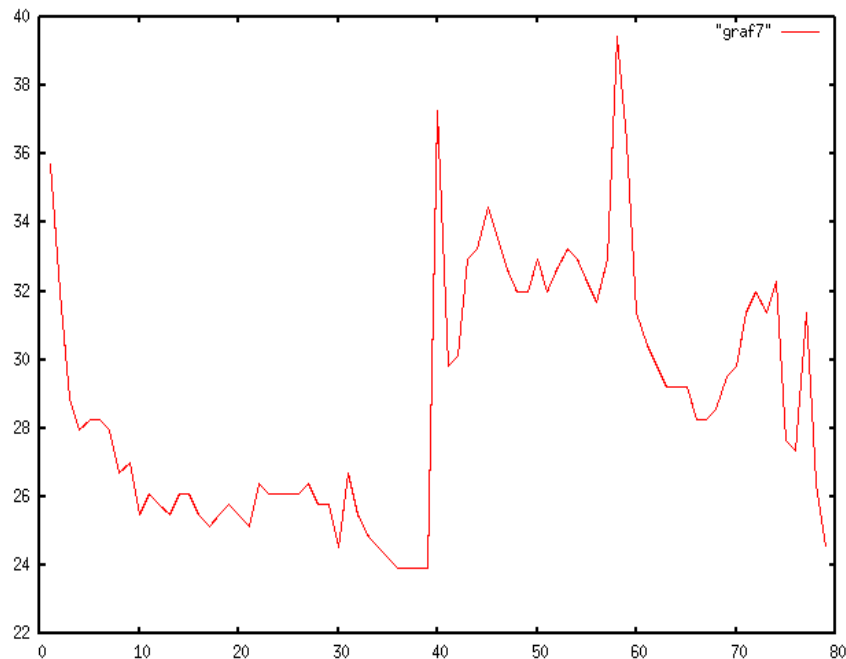


Imagen 58. Porcentaje de error (y) y nº de características (x) seleccionadas para características combinadas.

Aunque los experimentos probados no mejoran el *baseline*, sí que se obtienen buenos resultados mediante la selección de características. Sin embargo, debe ser complicado obtener una explicación a estos resultados ya que posiblemente se deben a un sobreentrenamiento. Para probar esta aproximación correctamente sería necesario disponer de una base de datos con un número mayor de mamografías que permitiera definir muestras diferentes de test, entrenamiento y validación de un tamaño adecuado. Aún así se comprueba que se puede mejorar la clasificación haciendo una elección de características posterior, sea cual sea la forma de obtener esas características.

4.5. Resultados con técnicas aplicadas sobre ventanas locales

Al ver que no se pueden mejorar mucho los resultados utilizando el histograma global de las mamografías por no disponer de información espacial, se ha pasado a un enfoque local del problema, dividiendo la imagen en ventanas de tamaño variable y aplicando diferentes técnicas para intentar obtener buenos resultados usando información local.

En una primera fase, se divide la mamografía en ventanas de tamaño W y se obtienen sus histogramas reescalados al rango de 0 a 255. Luego se utiliza PCA y kmeans para obtener particiones de C clusters. Por último, se clasifican según la cantidad de ventanas que tengan pertenecientes a cada cluster. Hay que señalar que empezaron a obtenerse resultados significativos cuando se normalizaron los histogramas locales. Sin este paso previo, la experimentación produce resultados similares a los de los experimentos anteriores (un error del 30-32%).

Se han realizado varios experimentos alterando el tamaño del histograma D (32, 64, 128, 256), de las ventanas W (40, 50, 60, 70, 80, 100), componentes principales NCP (2 a 20), número de clústers C (3 a 14) y varios vecinos k (para la fase local) y K (para la fase global)(10-15).

Ejemplo, con $D=32$, $W=50$, $C=9$ Y $k=K=10$:

NCP	Mejor error
3	27.33
5	27.64
7	26.40
9	27.64

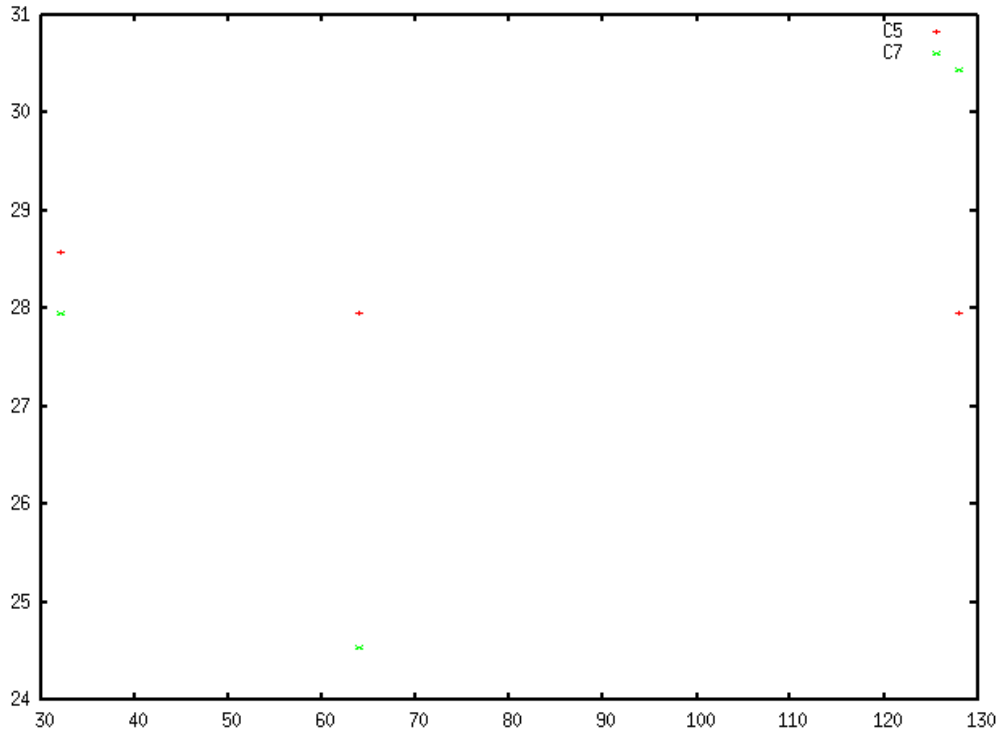


Imagen 59. Porcentaje de error y tamaño del histograma D , para $W=50$, $NCP=3$, $k=10$ y $K=10$.

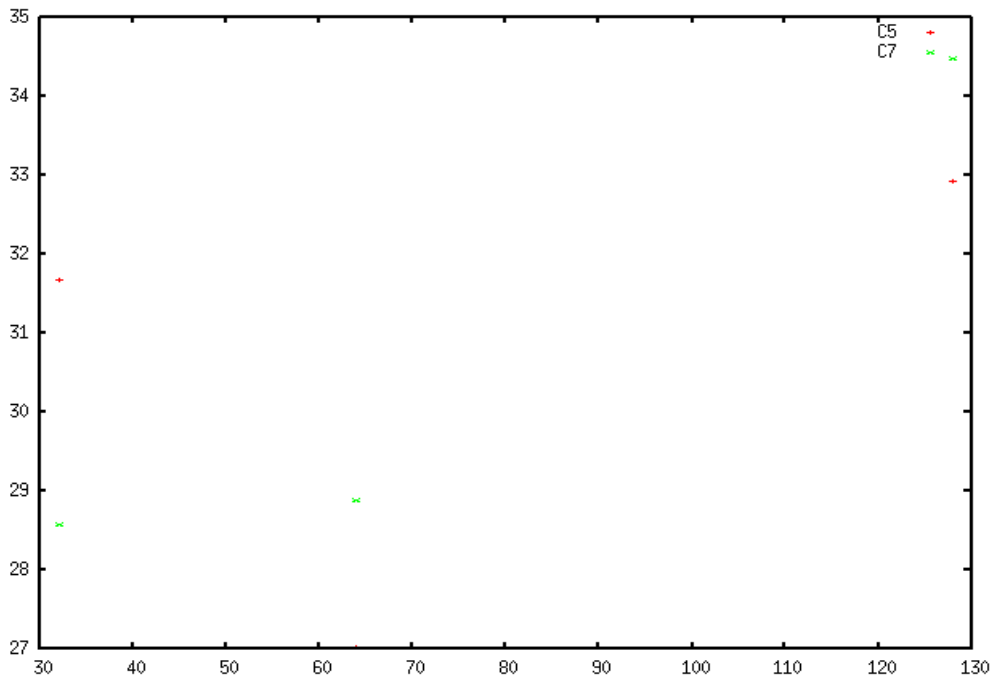


Imagen 60. Error y tamaño del histograma D , para $W=50$, $NCP=5$, $k=10$ y $K=10$.

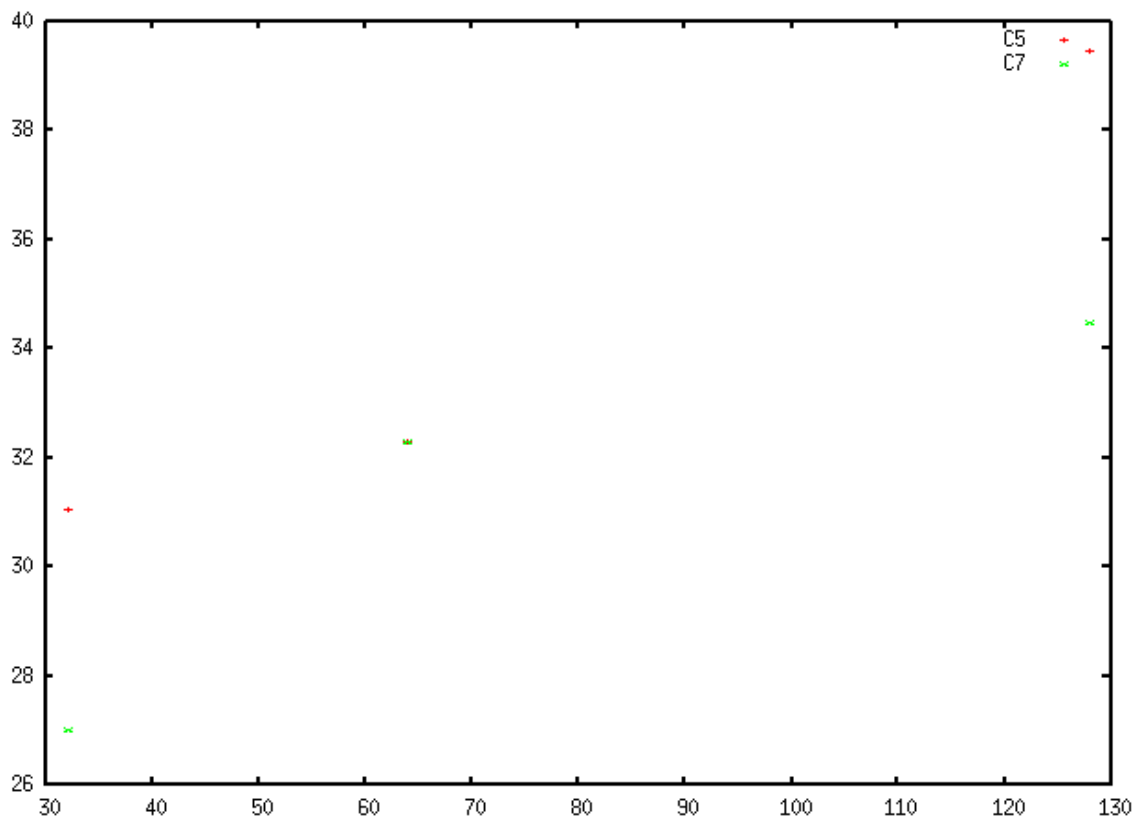


Imagen 61. Porcentaje de error y tamaño del histograma D , para $W=50$, $NCP=7$, $k=10$ y $K=10$.

En el artículo publicado [16] se presentaron los resultados alcanzados con un error del **22.04 %** de error, y se dejaba constancia del mínimo *Major Error* que se alcanzaba respecto a otros trabajos. El *Major Error* es la confusión que se da entre las clases extremas, representa el error “más grave” que puede darse por darse entre las clases más diferentes que en este caso es de sólo un 1.24% frente a otras técnicas en los que el mínimo hasta el momento era del 3.34%.

Fatty Glandular Dense			
Fatty	94	8	3
Glandular	11	67	19
Dense	1	29	90

Imagen 62. Matriz de confusión publicada en el artículo

Experimentaciones más minuciosas consiguieron incluso mejorar levemente el resultado, llegando a un **21.74 %** de error con $D=32$, $W=40$, $NCP=5$, $C=6$, $k=10$ y $K=9$, usando clasificación por votos.

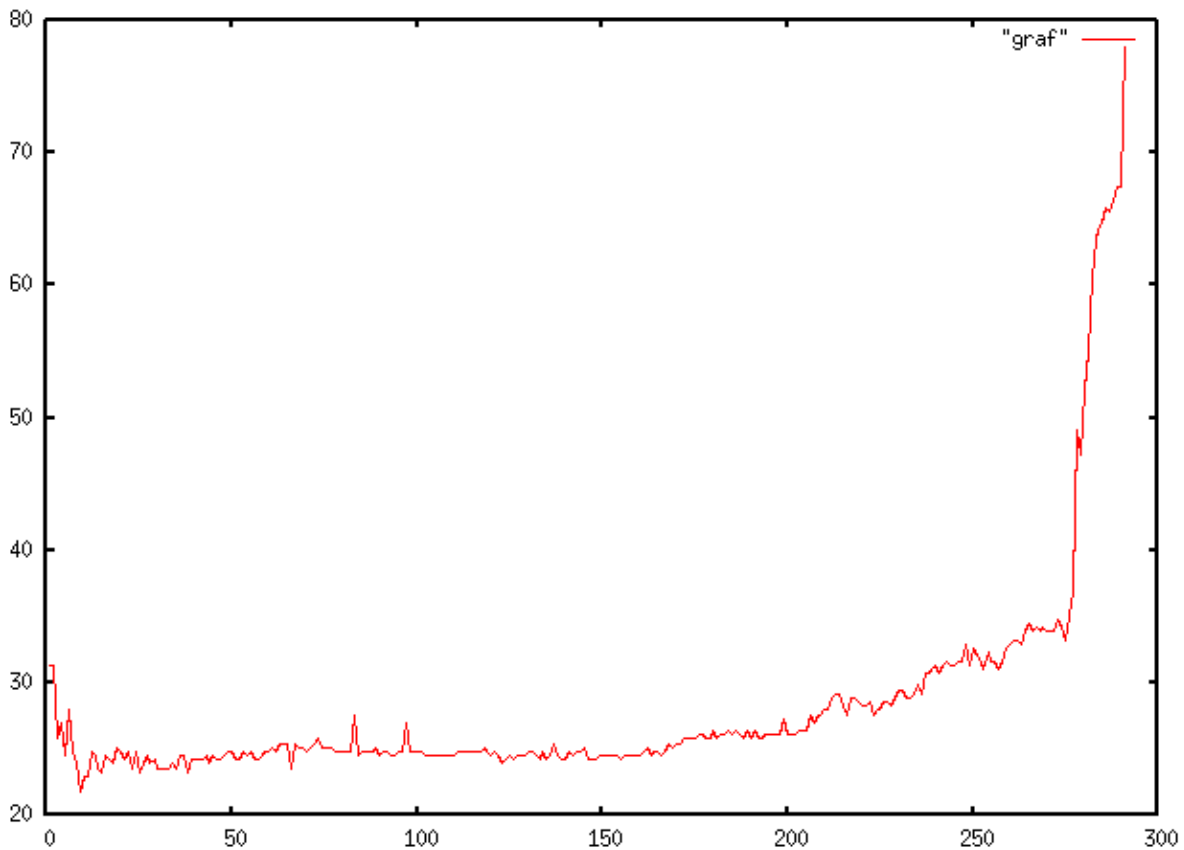


Imagen 63. Porcentaje de error y n1 de vecinos para $D=32$, $W=40$, $NCP=5$, $C=6$, $k=10$ y $K=9$.

Los centroides del k-means para este resultado se establecen en:

C0:	1780835.67	-104440.01	352177.71	66894.31	13149.86
C1:	-1326068.34	622163.98	318089.76	87182.42	62921.21
C2:	86059.84	1471088.38	173223.86	-115838.02	39985.31
C3:	-3104189.62	-2554250.83	-3053583.57	-495639.61	28064.68
C4:	1458280.88	1331217.21	-362551.10	-472914.08	118528.98
C5:	1149181.99	1781477.70	-718190.53	1083695.17	-162353.76

En vez de decidirse por un único clúster, se modificó el experimento para que se tuvieran en cuenta todos los clústers (en porcentaje) de modo que se disponga de información más precisa, evitando descartar clústeres importantes por elegir únicamente uno. Sin embargo, esta modificación no alteró de manera significativa los resultados obtenidos hasta ahora.

Se usaron otras características alejadas del histograma para ver los resultados. En este caso se calcularon los valores de media, varianza, asimetría $((1/N) * ((x-media/varianza)^3))$ y curtosis $((1/N) * ((x-media/varianza)^4 - 3))$ de cada subventana. Con estos operadores estadísticos se pretendían encontrar características útiles tales como deformaciones, asimetrías, semejanza a una distribución normal, etc. Sin embargo, los pobres resultados que se obtuvieron nos obligaron a buscar otras alternativas...

Se repitió el proceso de clustering que tan buenos resultados ha dado pero usando texturas obtenidas mediante matrices de coocurrencia en vez de usar el histograma normalizado. Repitiendo los parámetros del anterior experimento, no se consiguió llegar a porcentajes de error parecidos, manteniéndose el histograma normalizado local como el mejor hasta el momento.

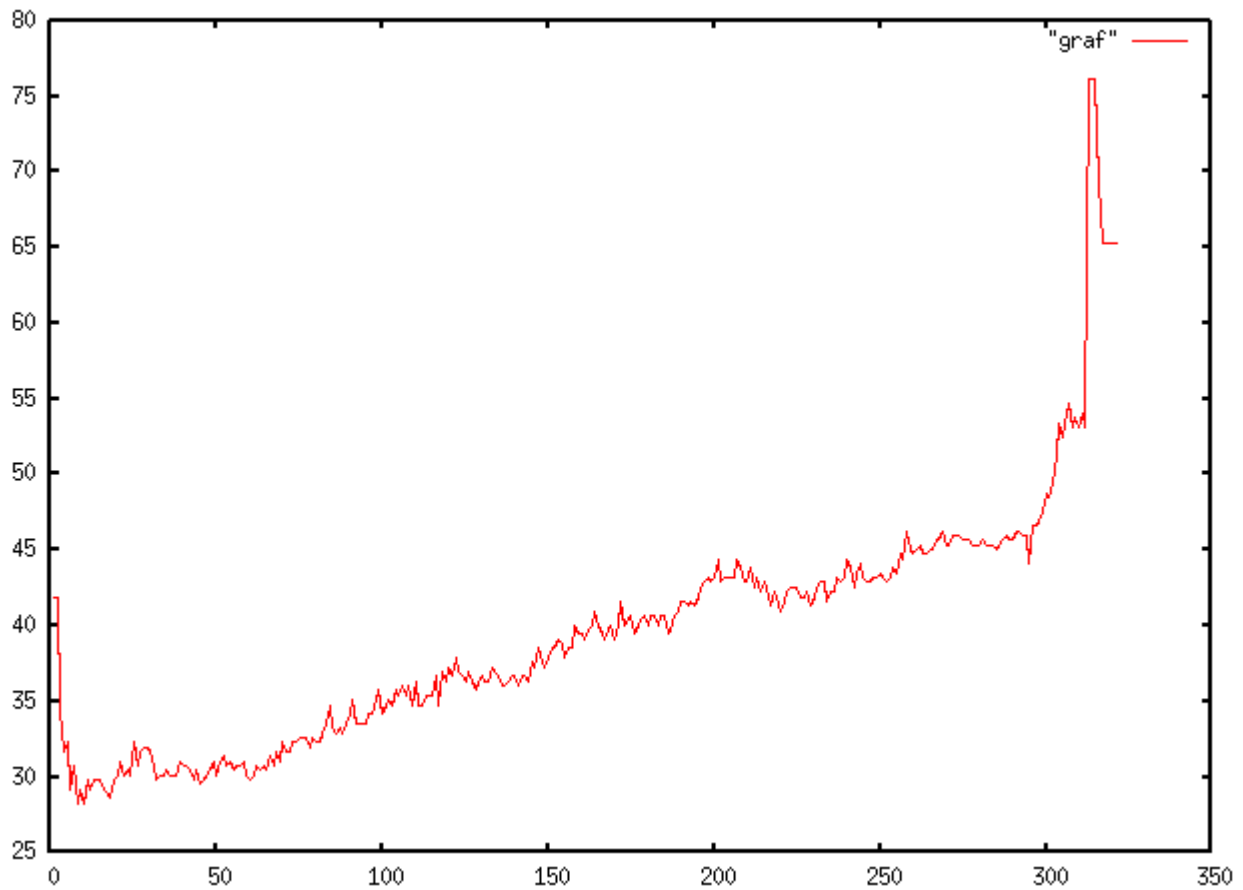


Imagen 64. Porcentajes de error y n° de vecinos para $D=64$, $W=60$, $NCP=3$, $C=6$ y $k=10$.

El mejor resultado con texturas se alcanzó con **28.26** % de error y parámetros $D=64$, $W=60$, $NCP=3$, $C=6$, $k=10$ y 10 vecinos.

Sin reescalar el histograma, pero aplicando previamente una reducción en la escala de grises y recalculando el histograma, se obtuvieron unos resultados más pobres que simplemente haciendo reescalado de histogramas...

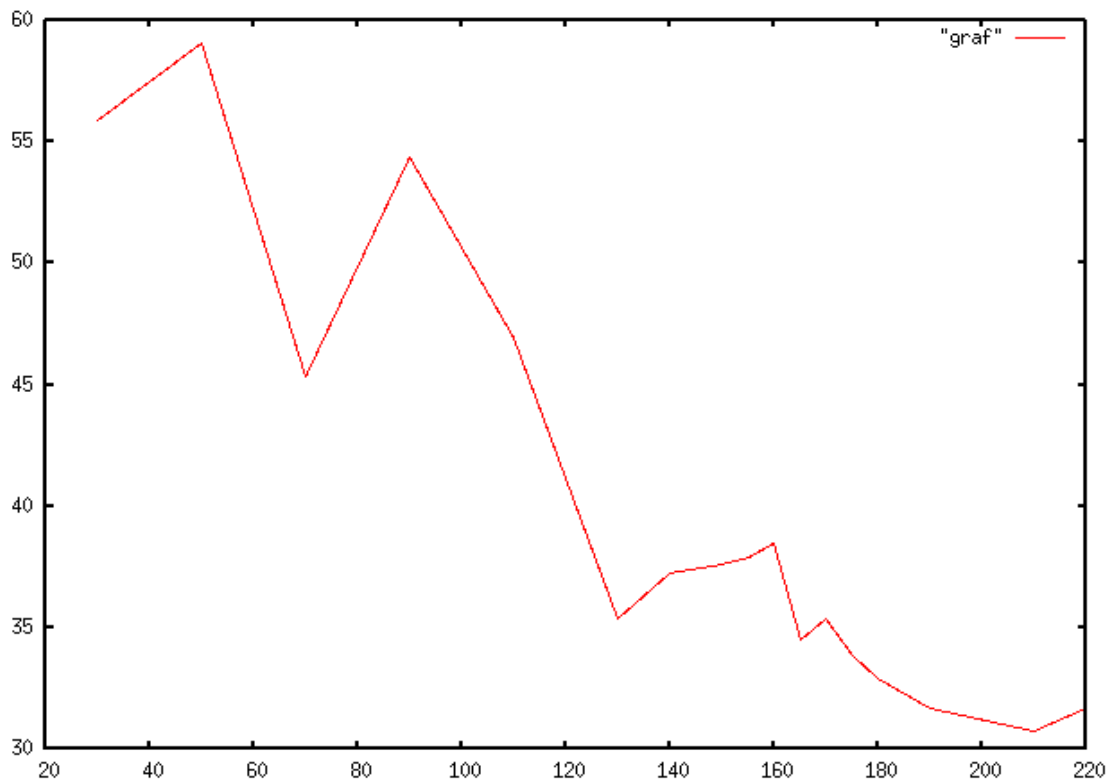


Imagen 65. Porcentaje de error y reescalados de colores para $W=60$, $NCP=3$, $C=5$ y $k=K=10$.

El mejor resultado en este caso fue de **30.74 %** de error con $W=60$, $NCP=3$, $C=5$, $k=10$, $K=10$, y un reescalado a 210 niveles de gris.

Posteriormente, se obtuvo una clasificación en cuatro clases de la base de datos MIAS, y se empleó el método descrito sobre la nueva clasificación para ver los resultados. En esta nueva clasificación hay 87 mamografías de la clase 1, 103 de la clase 2, 95 de la clase 3 y 37 pertenecientes a la clase 4.

Se realizó una batería de pruebas idéntica a la realizada con la clasificación en tres clases y se obtuvo como mejor resultado un error del **24.53 %**, con parámetros $D=64$, $W=60$, $NCP=7$, $C=7$, $k=10$ y 21 vecinos.

	1	2	3	4
1	73	12	1	0
2	14	81	27	0
3	0	10	65	13
4	0	0	2	24

Imagen 66. Matriz de confusión

El incremento del 2,81% de error es debido al aumento de número de clases. De nuevo, es importante destacar que no hay confusión entre las clases extremas (1 y 4) lo que le da bastante robustez al método propuesto.

Por último, se ha probado a clasificar un subconjunto de la base de datos DDSM de 4 clases y mayor complejidad con la técnica desarrollada.

	1	2	3	4
1	44	22	2	0
2	65	266	54	16
3	2	38	165	57
4	0	7	33	45

Imagen 67. Matriz de confusión DDSM

El error del 36% es explicable por los propios problemas de la DDSM. Se puede apreciar por la distribución de los datos del subconjunto elegido que no hay una muestra equitativa para cada clase: 68 para la clase 1, 401 para la clase 2, 262 para la clase 3 y 85 mamografías para la clase 4 para un total de 816 imágenes. A esto hay que añadirle la pérdida de información producida al convertir el formato LJPEG en uno válido, la presencia de imágenes digitalizadas con diferentes aparatos y a diferentes tamaños, etc. De todas formas puede apreciarse incluso en estas condiciones uno de los beneficios del método propuesto: la ausencia de confusión entre las clases extremas 1 y 4.

El método propuesto es el que mejor resultados ha dado con la base de datos MIAS de entre los estudiados, seguido de cerca por la selección de características con la que habría que utilizar un conjunto de imágenes mayor para asegurar los resultados.

Método	Porcentaje de error alcanzado
Histograma global normalizado con reducción	31.37%
Histograma global normalizado con PCA	31.37%
Multiresolution Histogram	35.09%
Multiresolution Histogram con PCA	30.34%
Histograma por Regiones	32.92%
Histograma por Regiones con PCA	30.12 %
Combinación Multiresolution y Por Regiones	36.65%
Selección de características	23.91%
Propuesto (Histogramas Locales)	21.74%

Imagen 68. Comparación métodos estudiados

4.6. Estudio gaussiano de la densidad

Se realizó también un estudio gaussiano de la densidad con la base de datos MIAS. En la metodología que se usa, de cada mamografía se extrae un número N de ventanas locales sobre la región de interés (la mama, eliminando el fondo y el músculo pectoral por no aportar información sobre la densidad de la mama) de un tamaño DxD y se clasifica con la etiqueta de la clase de la mamografía, formando así los vectores de características que se usarán posteriormente para clasificar la mama. Según se ha observado en otros experimentos con mamografías, la utilización de información ecualizada o no puede ser significativa así que se emplearán los dos tipos de ventanas locales para decidir qué preproceso es el más adecuado en esta

ocasión.

El clasificador gaussiano se basa en las distribuciones gaussianas. En estas distribuciones, la probabilidad se calcula mediante:

$$p(\mathbf{x}) = (2\pi)^{-\frac{D}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^t \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)$$

Donde Σ es la matriz de covarianza y $\boldsymbol{\mu}$ es la media. A la hora de crear el clasificador, se utilizará el clasificador de Bayes particularizado a las funciones de densidad gaussianas. El clasificador de Bayes sería:

$$c^*(\mathbf{x}) = \arg \max_c g_c(\mathbf{x}) \quad g_c(\mathbf{x}) = \mathbf{x}' W_c \mathbf{x} + \mathbf{w}_c^t \mathbf{x} + w_{c0}$$

donde, utilizando los valores de las gaussianas:

$$\begin{aligned} W_c &= -\frac{1}{2} \Sigma_c^{-1} \\ \mathbf{w}_c &= \Sigma_c^{-1} \boldsymbol{\mu}_c \\ w_{c0} &= \log p(c) - \frac{1}{2} \log |\Sigma_c| - \frac{1}{2} \boldsymbol{\mu}_c^t \Sigma_c^{-1} \boldsymbol{\mu}_c \end{aligned}$$

La versión con mixturas modifica la función de densidad de modo que sea una combinación lineal de componentes:

$$p(\mathbf{x} | \Theta) = \sum_{c=1}^C p_c p(\mathbf{x} | c, \Theta')$$

con los parámetros:

$$\Theta = (p_1, \dots, p_C; \Theta')^t$$

y desconocidos:

$$p_1, \dots, p_C \geq 0 \quad p_1 + \dots + p_C = 1$$

Para el caso de las gaussianas, sería:

$$p(\mathbf{x} | c, \Theta') = (2\pi)^{-\frac{D}{2}} |\Sigma_c|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_c)^t \Sigma_c^{-1} (\mathbf{x} - \boldsymbol{\mu}_c)\right]$$

con $c = [1 \dots C]$ y

$$\Theta' = (\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_C; \Sigma_1, \dots, \Sigma_C)^t$$

donde se encuentran todas las medias y matrices de covarianza.

El clasificador de mixturas se implementa utilizando una estrategia E.M. para

obtener la máxima verosimilitud para las componentes que formarán la mixtura.

En una primera fase, se implementará un clasificador gaussiano simple (con una componente por clase) con el que se clasificará cada vector de la mamografía correspondiente. Una vez clasificados todos los vectores de la mamografía, se procederá a un recuento de los vectores clasificados en cada clase y se clasificará la mamografía como la que corresponde a la clase con más vectores. Se establece, por lo tanto, una gaussiana por cada clase y se clasifica cada subventana de la imagen. El entrenamiento se hará utilizando cada subventana de las mamografías de la clase C como pertenecientes a la clase C.

Posteriormente, se implementará un clasificador de mixturas gaussianas, con el que se repetirá el experimento anterior. Las componentes de cada mixtura se aproximarán utilizando una estrategia EM, esperando obtener una mejora significativa al experimento anterior. En este caso, se establece una mixtura de gaussiana por cada clase, y cada subventana se clasifica entre cada mixtura para tratar de decidir a qué clase pertenece la imagen, variando en los experimentos el número de componentes de las mixturas. La importancia de la partición inicial de los datos se ha tenido en cuenta al repetir los experimentos tres veces con una inicialización aleatoria del conjunto inicial. Los resultados expuestos en las figuras corresponden a los mejores de las tres repeticiones. Aunque deberían realizarse más repeticiones, la falta de tiempo ha obligado a la restricción del número de repeticiones.

Por último, se sustituirá el clasificador de k-vecinos utilizado en el experimento complejo que dio buenos resultados (El del apartado anterior, que usa los histogramas locales) por el clasificador de mixturas gaussianas comparando los resultados con los obtenidos con k-vecinos. En este experimento, la clasificación se realiza en dos pasos: primero se utiliza k-means para dividir los histogramas de las subventanas (reducidos con PCA) en clusters y, después, se clasifica cada mamografía según la cantidad de subventanas que tenga pertenecientes a cada cluster. Será esta segunda clasificación la que se sustituirá por el clasificador de mixturas gaussianas (ver el apartado anterior para más detalles). De nuevo, se repiten los experimentos tres veces inicializando aleatoriamente las mixturas y se muestra el resultado mejor.

Los resultados utilizando un clasificador gaussiano simple muestra resultados pobres con las tres clases de la mamografía. La clase intermedia G se confunde totalmente entre las clases extremas D y F, de modo que el clasificador puede diferenciar entre las mamas más y menos densas (las de las clases extremas D y F) pero tiene problemas al decidir cuándo se encuentra en una densidad intermedia (G).

	D	G	F
D	106	0	6
G	80	0	24
F	33	0	73

	D	G	F
D	90	0	22
G	40	0	64
F	0	0	106

Imagen 69. Matrices de confusión para ventanas ecualizadas 8x8 y 16x16.

Se puede observar que el tamaño de la ventana influye en qué clase se confunde la clase G, ya que la mayoría se clasifica como D o F dependiendo del tamaño de la ventana. A partir de este punto se emplean únicamente las ventanas de la imagen ecualizadas localmente, ya que usar las ventanas sin ecualizar presenta unos resultados peores.

	D	G	F
D	74	0	38
G	30	0	74
F	13	0	93

Imagen 70. Matriz de confusión para ventanas 16x16 sin ecualizar.

El uso de varios componentes por gaussiana permitió al clasificador identificar mamografías de la clase intermedia G (que anteriormente no conseguía encontrar), pero al precio de confundir las clases extremas más que antes. Otro problema que se presenta es la lenta mejora que se experimenta al aumentar el número de componentes gaussianos, y que clasifica pocas mamografías como mamografías de clase G:

	D	G	F
D	38	40	34
G	42	38	24
F	2	10	94

	D	G	F
D	46	42	24
G	52	28	24
F	6	6	94

	D	G	F
D	53	35	24
G	50	40	14
F	8	4	94

Imagen 71. Evolución de la matriz de confusión en ventanas 8x8, aumentando el n° de componentes de 20 en 20, empezando con 10 componentes.

Sin embargo, con este método se consiguió un error superior al 30% como mejor resultado, usando ventanas de 16x16 y utilizando 40 componentes. La confusión entre las clases G y D no parece esclarecerse utilizando mixturas, proporcionando un error demasiado elevado para ser significativo.

Se decidió modificar el número de componentes de la clase G, la más problemática, dejando con una componente a las clases D y F para ver si se podía obtener alguna mejora. Sin embargo, las pruebas que se realizaron con 10, 30 y 50 componentes para la clase G devolvían resultados similares a los primeros experimentos de gaussianas sin componentes y se abandonó la idea.

	D	G	F
D	104	0	8
G	96	0	8
F	38	0	68

Imagen 72. Matriz de confusión para un clasificador con 10 componentes en la clase G y sólo uno en las clases D y F.

El uso de las mezclas gaussianas para clasificar las subventanas reducidas en clusters (la tercera parte del trabajo) sí que obtuvo buenos resultados.

	D	G	F
D	71	39	2
G	16	73	15
F	3	22	81

Error: 30.1%

	D	G	F
D	70	37	2
G	17	71	16
F	3	14	89

Error: 29.2 %

	D	G	F
D	73	37	2
G	17	71	16
F	3	14	89

Error: 27.6 %

Imagen 73. Matrices de confusión para ventanas 32x32, reducción PCA a 3 componentes y 5 clusters diferentes.

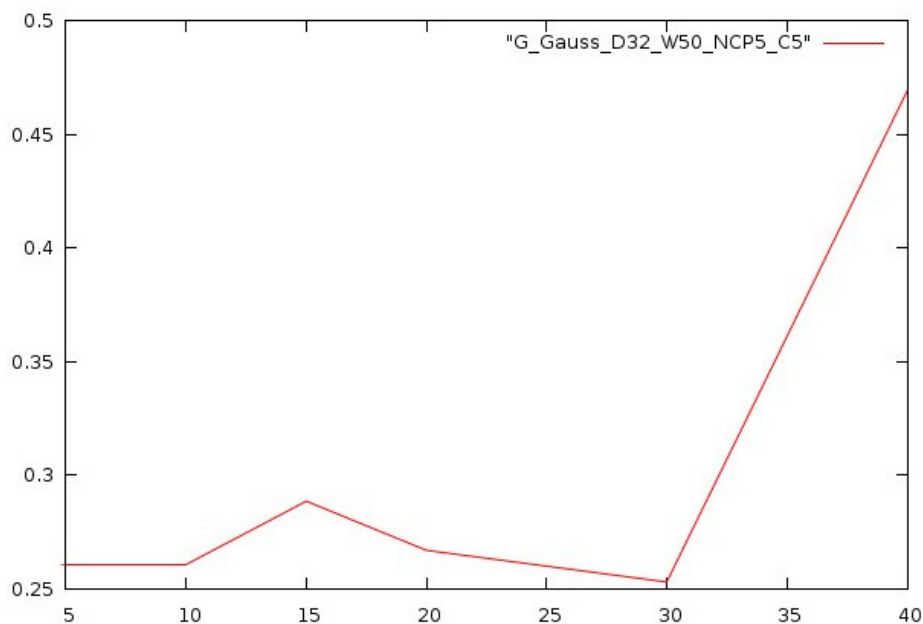


Imagen 74. Evolución del error según el número de componentes en las mezclas, para ventanas de 50x50 reducidas a 5 componentes y clasificadas en 5 clusters.

En esta ocasión, se aprecia que los buenos resultados se aproximan a los obtenidos mediante k-vecinos aunque no consiguen superarlos, ya que el mejor resultado obtenido con k-vecinos ha sido del 22.64% utilizando menos de 30 vecinos. La clasificación mediante mixturas ha conseguido un error del 26.0 % optimizando los parámetros y utilizando 30 componentes.

Se ha podido observar mediante los experimentos realizados que la clasificación gaussiana de las subventanas de las mamografías no es lo suficientemente potente como para conseguir diferenciar entre las tres clases. La aplicación de mixturas, aunque permite “encontrar” mamografías de la clase intermedia, sigue sin ser capaz de obtener resultados significativos por sí misma. Esto es debido a la clasificación inicial forzada de considerar cada subventana del mismo tipo que la mamografía a la que pertenece. Esta aproximación hace que la clase intermedia acabe confundida entre las más extremas, ya que al ser la clase medianamente densa contiene subventanas que podrían considerarse como densas o no densas (las que existirán en gran mayoría entre las clases extremas). Puede comprobarse mirando el recuento de ventanas de las mamografías de clase G, en las que las subventanas clasificadas son minoría (un 10-20%) respecto a las clasificadas como D o F (que se reparten el 90-80% restante).

El uso de mixturas, aunque permite distribuir mejor el espacio entre las tres clases, se encuentra de nuevo con el problema de tener clasificadas inicialmente las subventanas según la clase a la que pertenece la mamografía, con lo que se obliga al clasificador a tener ventanas similares como pertenecientes a diferentes clases. La mayor confusión a la que están sometidos los resultados en esta clasificación se deben a la presencia de subventanas similares clasificadas como pertenecientes a diferentes clases, haciendo la decisión más problemática.

Por último, al hacer una clasificación inicial mediante k-means se independizan las subventanas de la clase a la que pertenece la mamografía. Y al clasificar según los clusters a los que pertenecen cada subventana, se obtiene una clasificación más natural (y no tan forzada como en los casos anteriores). Y es en esta ocasión donde se aprecia que los resultados con mixturas se aproximan a los resultados obtenidos con k-vecinos, aunque no consiguen superarlos.

Con esto se pone de manifiesto la principal desventaja que tiene el abordar la estimación de la densidad en mamografías de manera automática mediante una estrategia de ventanas locales: el no poder determinar automáticamente si un píxel (o grupo de píxeles) es de tipo graso o denso.

5. Conclusiones

En este trabajo se ha hecho un repaso al estado actual del análisis de la densidad de las mamografías. Se han probado algunos métodos importantes extraídos directamente de la bibliografía para comprobar su funcionamiento y eficiencia y, por último, se ha presentado un nuevo método para la clasificación de la densidad que iguala a los mejores métodos en algunos aspectos y los mejora en otros. El lector habrá comprendido que la clasificación de mamografías en función de su densidad no es una tarea sencilla y a veces incluso frustrante, pero es importante trabajar en un problema tan actual como éste.

Se ha comprobado que la aplicación de una metodología de ventanas locales es superior a los métodos publicados en los que se emplea únicamente el histograma global. De este modo se puede afirmar que la clasificación de la densidad es un problema que no puede resolverse completamente utilizando únicamente el histograma global de la mamografía, ya que es incapaz de retener información espacial significativa para determinar la densidad. Con el uso de los histogramas a partir de las ventanas locales se intenta establecer una relación de “densidad” entre ventanas, permitiendo la clasificación de la mamografía en varios niveles de “densidad” y obteniendo una tasa de aciertos mayor que si se usa únicamente el histograma global.

La obtención de resultados similares con 3 o 4 clases en las imágenes de la base de datos MIAS muestra la robustez del método expuesto, mientras que los problemas observados al utilizar la base de datos DDSM pueden deberse más a la propia DDSM que a la experimentación. La pérdida de precisión al pasar del formato LJPEG a otro formato de imagen analizable, el uso de cuatro tipos distintos de digitalizadora y la diversidad de tamaños son problemas que han de tenerse siempre en cuenta a la hora de utilizar imágenes de la DDSM, unidos a la siempre presente subjetividad del experto que etiqueta las radiografías.

También se ha iniciado una metodología hasta ahora inédita en la literatura, como es la aplicación de las gaussianas a mamografías para establecer su densidad, que puede dar lugar a nuevos estudios que permitan mejorar la clasificación o enfocar el problema desde otro punto de vista. Existen varias técnicas de obtención de vectores de características usando el histograma global de las mamografías que obtiene resultados interesantes (de un error del 28-30%) con Support Vectors Machines o k-vecinos que pueden ser revisitados con un clasificador de mixturas gaussianas, aunque para la realización de este trabajo sólo se han explorado las posibilidades con ventanas locales. También puede intentar abordarse el problema de tratar de mejorar la confusión existente en el uso de mixturas sobre las ventanas locales directamente, tal vez mediante reducciones PCA con la esperanza de que se minimicen las ambigüedades entre subventanas similares de diferentes mamografías. Esta opción es la que espero tratar en un futuro, ya que aunque los resultados no han sido espectaculares resulta notable que se haya podido clasificar correctamente un 70% de la base de datos utilizando ventanas locales sin nada más. Todas estas interesantes posibilidades espero poder probarlas en breve, para ver si se consigue mejorar en algo los resultados actuales o para abrir nuevas posibilidades de investigación.

Método	Porcentaje de error alcanzado
Histograma global normalizado con reducción	31.37%
Histograma global normalizado con PCA	31.37%
Multiresolution Histogram	35.09%
Multiresolution Histogram con PCA	30.34%

Histograma por Regiones	32.92%
Histograma por Regiones con PCA	30.12 %
Combinación Multiresolution y Por Regiones	36.65%
Selección de características	23.91%
Propuesto (Histogramas Locales)	21.74%
Clasificador de mixturas gaussianas	30.00 %

Imagen 75. Comparación de métodos estudiados.

Se seguirá trabajando en el estudio de la densidad de mama en parte por el nuevo proyecto asignado por la Generalitat. Con este nuevo proyecto se dispondrá de una cantidad ingente de mamografías con la que poder comprobar la fiabilidad del selector de características y, además, añadir las características del método propuesto a las ya disponibles, además de disponer de varias clasificaciones proporcionadas por varios expertos para poder compararlas de manera más objetiva con los resultados que obtenga el ordenador.

También se intentarán nuevos métodos de aproximación al problema como, por ejemplo, la implementación de un sistema de *Cumulus* “automático, en el que el ordenador intentará encontrar un umbral (nivel de gris) a partir del cual se considerará tejido denso y clasificará la mamografía en función de los píxeles densos que encuentre.

Va a estudiarse la posibilidad de, sin dejar de lado el tema de la densidad, empezar a tratar la detección de zonas cancerosas o anómalas en la mama. Esto significará un estudio más detallado de la imagen en el que ya no importará su densidad sino las estructuras anómalas que se encuentren en su interior.

6. Apéndices

6.1. Base de datos MIAS

Nombrada con las siglas de *Mammography Image Analyse Society* (que es una asociación inglesa de investigación relacionada con las mamografías), es una base de datos de mamografías cuidadosamente seleccionadas del programa nacional británico de análisis de mamas. Las mamografías se han digitalizado mediante un escáner *Joyce-Lobel* a una resolución de 50 x 50 μm . La base de datos contiene imágenes de 1024x1024 de las mamas izquierda y derecha de 161 pacientes, en total está formada por 322 imágenes divididas en tres clases: 208 son normales, 63 con cáncer benigno y 51 con cáncer maligno. La base de datos también dispone de un fichero en la que se describe el análisis realizado por un radiólogo experimentado. Para cada imagen, se indica el tipo, localización, escala, densidad y otra información útil sobre ella realizada por el experto. También viene con una clasificación de densidad en tres clases: *granular (G)*, *fatty (F)* y *Dense (D)*. En esta clasificación las mamografías más densas corresponden a la clase D y las menos densas a la clase F, estando la clase G en una situación intermedia entre las clases extremas. En esta distribución hay 104 mamografías G, 112 D y 106 de la clase F.

Es una base de datos que puede obtenerse de forma gratuita por Internet y es ampliamente utilizada en los estudios sobre radiografías de manera que puedan compararse resultados. Su principal inconveniente es la clasificación en tres clases de densidad, que es poco realista, y el hecho de que realmente hay 161 imágenes “diferentes” ya que las radiografías izquierda y derecha de una paciente son muy similares en cuanto a su densidad. Si no es tenido en cuenta a la hora de realizar experimentos con esta base de datos, los resultados pueden no ser significativos.

6.2. Base de datos DDSM

La *Digital Database for Screening Mammography* es una base de datos de mamografías gratuita al alcance de la comunidad investigadora. Este conjunto está formado por aproximadamente 2500 estudios, cada uno con dos imágenes de cada mama e información relevante sobre el paciente (edad, densidad de mama, etc.) y sobre las imágenes (digitalización, resolución, etc.).

La DDSM está organizada en casos y volúmenes. Un caso es una colección de imágenes e información que corresponde a un examen mamográfico de un paciente. Un volumen es sólo una colección de casos ordenados para una facilidad de distribución. Cada caso está formado por entre 6 y 10 ficheros que pueden ser o ficheros de texto de información con extensión .ics o imágenes. Hay una imagen en pgm que contiene una vista previa de las mamografías, y las mamografías están en formato ljpeg (formato jpg sin pérdidas).

El principal inconveniente de esta base de datos es el formato en el que se encuentran las mamografías. El formato ljpeg es un formato obsoleto y sin soporte, lo que dificulta el análisis de las imágenes. En el siguiente apartado se indica el programa utilizado para el paso de las imágenes ljpeg a formatos más manejables como pgm o bmp. Otro de los problemas que presenta este conjunto de imágenes es que en los estudios en los que se usa no indican el subconjunto utilizado, con lo que se hace difícil comparar resultados al no saber qué conjuntos han sido utilizados. También se ha de indicar que las imágenes no tienen el mismo tamaño, lo que añade complejidad a la hora de analizar las mismas. Por

otra parte, la clasificación en 4 clases es mucho más realista y cercana a la verdad.

6.3. Recursos utilizados

Aunque he escrito (muchísimos) scripts para *Linux* y programas en C para la realización de estos experimentos y pruebas, indico a continuación los programas que he usado o modificado para mi conveniencia, y dar las gracias a sus creadores:

- Librería *libitknn* del Instituto Tecnológico de Informática (ITI): k-vecinos y kdtrees.
- Librería *libitidip* del Instituto Tecnológico de Informática (ITI): carga de imágenes.
- *SVM-Struct*: Clasificador de *Support Vector Machines*.
- Varios *scripts* del profesor Rafa Llobet.
- Librería *ljpg*: paso de *ljpg* a formatos de imagen válidos. Obsoleta pero vital para crear un programa que pase de un formato a otro.

7. Bibliografía

-Artículos:

- E.J. Feuer, L.M. Wun: DEVCAN: Probability of Developing or Dying of Cancer. Version 4.0. Bethesda MD: National Cancer Institute. (1999) [1]
- D.B. Kopans: The positive predictive value of mammography. *AJR*, Vol. 158 (1992) 521–526 [2]
- G.M. te Brake, N. Karssemeijer: Automated detection of breast carcinomas that were not detected in a screening program. *Radiology*, Vol. 207 (1998) 465–471 [3]
- M. Wallis, M. Walsh et al.: A review of false negative mammography in a symptomatic population. *Clin Radiol* Vol. 44 (1991) 13–15 [4]
- J.N. Wolfe: Breast pattern as an index of risk for developing breast cancer. *AJR*, Vol. 126 (1976) 1130–1139 [5]
- J.N. Wolfe: Risk for breast cancer development determined by mammographic parenchymal pattern. *Cancer*, Vol. 37 (1976) 2486–2492 [6]
- N.F. Boyd, J.W. Byng, R.A. Jong, et al.: Quantitative classification of mammographic densities and breast cancer risk: Results from the Canadian national breast screening study. *J. Nat. Cancer Inst.*, Vol. 87 (1995) 670–675 [7]
- A.F. Saftlas, R.N. Hoover, L.A. Brinton, et al.: Mammographic densities and risk of breast cancer. *Cancer*, Vol. 67 (1991) 2833–2838 [8]
- American College of Radiology (ACR): Illustrated Breast Imaging Reporting and Data System (BI-RADS). 3rd edn. Reston, VA: American College of Radiology, (1998) 167–181/ [9]
- N. Jamal, K.H. Ng, L.M. Looi, et al.: Quantitative assessment of breast density from digitized mammograms into Tabar's patterns. *Phys. Med. Biol.*, Vol. 51 (2006) 5843–5857 [10]
- N. Karssemeijer: Automated classification of parenchymal patterns in mammograms. *Physics in Medicine and Biology*, Vol. 43 (1998) 365–378 [11]
- P.K. Saha, J.K. Udupa, E.F. Conant, D. Sullivan: Breast tissue density quantification via digitized mammograms. *IEEE Trans. on Medical Imaging*, (8) Vol. 20 (2001) 792–803 [12]
- A. Oliver, J. Freixenet, A. Bosch, et al.: Automatic classification of breast tissue. *Lecture Notes in Computer Science*, Vol. 3523 (2005) 431–438 [13]
- A. Oliver, J. Freixenet, R. Martí, et al.: A novel breast tissue density classification methodology. *IEEE Trans Inf Technol Biomed.*, Vol. 12 (2008) 55–65 [14]
- I. Muhimmah, R. Zwiiggelaar: Mammographic density classification using multiresolution histogram information. *Proceedings of the International Special Topic Conference on Information Technology in Biomedicine*, (2006) [15]
- R. Llobet, J.A. Solves: Mammographic density classification based on local histograms (2008) [16]
- R. Llobet, F. Luís-Perales, J.A. Solves: Estimación de la densidad mamaria asistida por ordenador (2008) [17]

-Página web de la Federación Española de Cáncer de Mama, <http://www.fecma.org/>

-Página web del Instituto Nacional del Cáncer, <http://www.cancer.gov/cancerinfo>

-Página web de la Wikipedia, <http://es.wikipedia.org/>

-Apuntes *Informática Médica*; M. Robles, S. Díez y J.V. Manjón.

-Apuntes *Biometría*; R. Paredes.

-Apuntes *Introducción al reconocimiento de formas*; A. Juan.