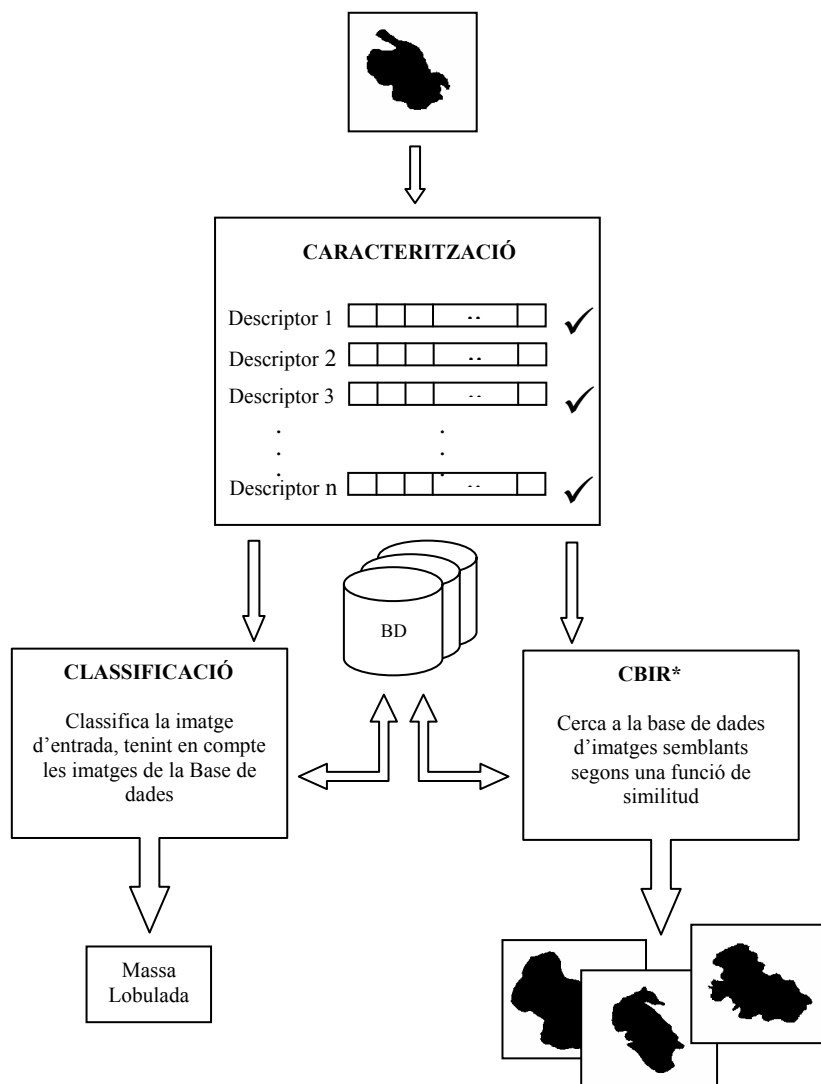


## 1 ) INTRODUCCIÓ

El present Projecte Final de Carrera (PFC) s'emmarca dins el projecte HRIMAC (Herramienta de Recuperación de Imágenes Mamográficas por Análisis de Contenido), iniciat l'any 2003 i subvencionat pel "Ministerio de Ciencia y Tecnología" i el fons FEDER. En aquest projecte hi participa la Universitat de Girona, la Universitat Ramón Llull, i especialistes de l'hospital de Girona Josep Trueta. La planificació del projecte es realitza a tres anys vista, distribuint les tasques a realitzar entre diferents directors responsables dels tres centres.

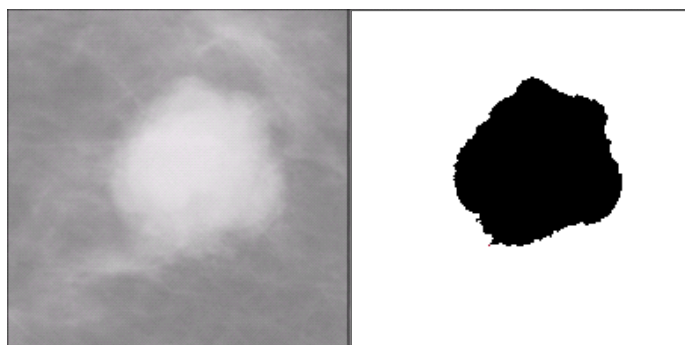
El projecte HRIMAC, pretén ser un projecte de recuperació d'imatges mamogràfiques per contingut. Així doncs, a partir d'una imatge mamogràfica sobre la qual es pretén emetre un diagnòstic, HRIMAC busca a la base de dades les mamografies més similars. D'aquesta manera, cada cerca proporciona un conjunt limitat de casos amb característiques molt semblants a la mamografia que s'està estudiant. L'anàlisi d'aquests casos, pot ajudar al radiòleg a diagnosticar amb més garanties d'èxit i augmentar d'aquesta forma el grau d'eficàcia en la interpretació.



**Figura 1.-** Esquema bàsic de l'aplicació desenvolupada en aquest PFC

## 2) DESCRIPCIÓ I OBJECTIUS

En aquest PFC s'han estudiat, discutit, analitzat i implementat una petita part del projecte HRIMAC, concretament la que es correspon a la caracterització de lesions segons la seva forma. Si bé, abans d'aplicar els algorismes per extreure'n la informació que ens interessa, cal un procés previ de segmentació per tal de diferenciar clarament el que és la lesió de la resta de la imatge. Així doncs, en aquest PFC s'ha treballat amb imatges com la que es mostra a la figura 2.b, imatges que contenen masses ja segmentades. A partir d'aquestes imatges, s'han estudiat diferents descriptors de forma per tal de determinar quins són els millors a l'hora de tractar amb lesions mamogràfiques. La figura 1 mostra esquemàticament les tasques dutes a terme.



**Figura 2.- a) Mamografia amb una massa b) mamografia segmentada**

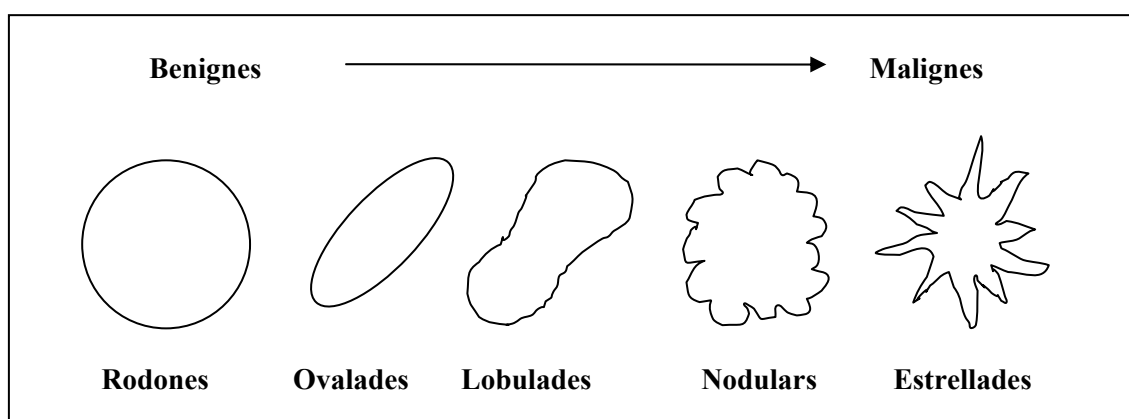
Concretament, a partir d'una imatge mamogràfica que conté una lesió segmentada, ens hem centrat en:

- Estudiar diferents mètodes tradicionals d'extracció de forma (àrea, perímetre, elongació...) que ens puguin donar una indicació de la mida, amplada, o posició del tumor.
- Analitzar i implementar diferents mètodes de descripció de forma que es basen en el contorn.
- Testejar els mètodes implementats tant per imatges sintètiques com per imatges mamogràfiques:
  - En una aplicació CBIR, de recuperació d'imatges per contingut.
  - En una aplicació de classificació, de reconeixement de casos patològics.
- Crear una base de dades independent de la base de dades HRIMAC per tal de poder testejar els descriptors de forma.
- Desenvolupar una aplicació prou flexible que permeti la incorporació de nous mètodes d'extracció de característiques, ja sigui de forma o de textura.

### 2.1) Quina importància té la forma?

Hi ha diferents lesions que es poden trobar en una imatge mamogràfica: microcalcificacions, masses i masses espiculars. El tipus de lesions que han estat objectiu d'estudi d'aquest PFC són les masses, i com a tal anem a detallar com la seva forma és reveladora de la seva nocivitat. Tant és així, que aproximadament entre el 80 i el 85% de càncers de pit són diagnosticats per l'aparença del tumor a la mamografia. Les principals característiques que s'utilitzen per diagnosticar la nocivitat d'un tumor són densitat, tamany, forma i contorns.

Pel que fa a les masses que es poden trobar en una mamografia, les que són d'origen maligne, sovint s'escampen pel voltant del teixit. Aquest fet provoca que el perfil consisteixi en fils fins extesos irregularment des del centre de la massa cap a fora. Aquest tret diferenciador de les masses malignes fa que poguem utilitzar la forma de la massa per discriminar entre masses malignes i masses benignes. Tot i així, cal deixar clar que la nostra intenció no és diagnosticar, sinó exclusivament utilitzar la forma per recuperar casos de la base de dades per tal que siguin de gran ajuda sobretot en mamografies poc clares. La Figura 3, mostra la gama de masses que més freqüentment podem trobar a les mamografies.



**Figura 3. – Esquema de les possibles masses presents a una mamografia, ordenades segons la seva benignitat**

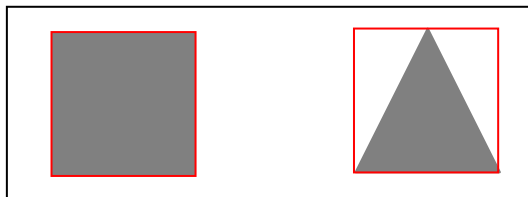
Donada una imatge els descriptors ens permetran obtenir informació per tal de caracteritzar la lesió o objecte que aquesta contingui. En altres paraules, ens permeten descriure la forma mitjançant un conjunt de valors numèrics, amb els quals podrem treballar. A continuació doncs, es fa una breu descripció dels descriptors de forma estudiats en aquest PFC.

### **3) DESCRIPTORS DE FORMA**

S'han tingut en compte dos tipus de descriptors: els descriptors bàsics de forma i els descriptors avançats de forma. Els descriptors bàsics de forma són de gran interès pel fet que si els utilitzem com a complement dels descriptors avançats, permeten fer cerques més restrictives o d'un interès més específic per part del metge. A més a més, ocupen poc espai i són ràpids de calcular.

#### **3.1) Descriptors bàsics de forma**

Els descriptors bàsics que s'han tingut en compte són: Àrea, Perímetre, Compactesa, Rectangle Mínim Envolvent, Alçada, Amplada, Elongació, Centre Gravetat. Amb aquests descriptors es podrà saber el tamany, la posició, etc, de l'objecte. És difícil però, determinar amb exactitud quina és la forma exacte que s'està analitzant principalment per dues raons. Podem tenir dos objectes amb diferent forma i en canvi tenir els descriptors bàsics iguals. Per exemple la figura 4 mostra dos objectes diferents però que tenen el mateix mínim rectangle envolvent i per tant també presentaran exactament els mateixos descriptors bàsics que s'extreuen a partir d'ell iguals.



**Figura 4. – Dues figures diferents que presenten el mateix mínim rectangel envolvent**

L'altre inconvenient és el fet que la majoria d'aquests descriptors donen valors diferents quan una mateixa figura es presenta amb diferent tamany o rotada. La figura 4 mostra com una rotació ens varia pràcticament tots els descriptors bàsics excepte l'àrea i el perímetre. La seva principal virtut és que són molt ràpids i que ens proporcionen una primera impressió de la figura que s'analitza. Aquesta primera impressió pot ser d'utilitat en cas de voler distingir entre figures molt i molt diferents o que essencialment es diferenciïn per alguna de les mesures que miren aquests descriptors.

### **3.2) Descriptors avançats de forma**

Per obtenir una representació que ens sigui d'utilitat, els descriptors avançats es basaran amb la informació del contorn. A partir de punts del contorn calcularem uns valors que anomenarem descriptors que ens caracteritzaran la figura. El que es pretén amb els descriptors és que el seu valor coincideixi per totes les figures que presentin la mateixa forma. D'aquesta manera un mateix objecte de la ha de tenir els mateixos descriptors independentment del tamany, la posició o de si està rotat.

Entre les diferents propostes, s'han considerat el mètode de Chang i el mètodes basats en Fourier:

- *Mètode de Chang.* Descriu la figura a partir de punts característics del contorn. Col·loquialment parlant, es busquen les cantonades de la figura per fer-ne una representació que ens sigui d'utilitat.
- *Mètodes de Fourier.* A la bibliografia hi trobem diferents propostes que es basen en la representació de figures mitjançant la representació espectral del contorn que s'obté com a resultat d'aplicar la transformada de Fourier. Existeix un espai, el domini freqüencial, on és possible representar figures i fer que la comparació sigui una cosa directe. El que fa possible que la comparació resulti directe és la normalització que es pot aplicar als descriptors fent que una figura amb diferent tamany, i/o diferent rotació, i/o diferent posició, tinguin una mateixa representació. S'han estudiat tres mètodes per tal d'obtenir la forma d'un objecte basant-nos en aquests descriptors.

## **4) RESULTATS**

Per tal d'analitzar els mètodes anteriorment comentats s'han desenvolupat dos sistemes: un sistema CBIR (Content Based Image Retrieval) i un sistema de classificació. Aquests dos s'han testejat amb imatges que contenen lesions mamogràfiques, amb imatges geomètriques (triangles, quadrats), i amb imatges sintètiques prèviament binaritzades (cotxes, ànecs etc..).

Es pot dir que els descriptors obtinguts amb el mètode de Chang no són una bona eina per tal de descriure la forma dels objectes, ja que aquests depenen del nombre de punts de

màxima corbatura que es detectin a l'objecte. I aquests punts no sempre són trobats correctament. Cal notar que no hi ha un mètode clarament millor. De manera general els millors resultats s'obtenen aplicant coordenada complexa. Tot i així la distància al centre té uns resultats aproximats i té l'aventatge que treballa amb la meitat d'informació. Potser seria interessant treballar amb coordenada complexa per l'eficiència en velocitat que això suposaria. Tot i la millora que poguéssim obtenir també és cert que els resultats de fer-ho amb coordenada complexa han estat pràcticament idèntics per les mamografies, donant molt bons resultats el descriptor únic, que fins llavors no havia funcionat gaire bé.

### **5) CONCLUSIONS I TREBALLS FUTURS**

Durant aquest projecte s'ha desenvolupat una aplicació específica per analitzar descriptors de forma. S'ha desenvolupat una aplicació on la integració de nous mètodes d'extracció de característiques (ja siguin de forma o textura), es pogués realitzar de forma senzilla. Podem dir que els objectius marcats a l'inici del PFC han estat complerts. I més que proporcionar un mètode millor que els altres seria interessant afegir a HRIMAC la combinació del descriptor únic amb la distància al centre per ser els més ràpids i que millors resultats donen.

Entre altres, podríem considerar els següents treballs futurs:

- Extreure el mínim rectangle envolvent, tenint en compte la orientació de la figura.
- Intentar treballar amb els N primers descriptors de Fourier, i no amb tots. D'aquesta manera, ens perdríem detalls de la figura, però la forma genèrica ja quedaria captada.
- Buscar altres signatures que permetin agrupar els descriptors en un descriptor únic, intentant buscar un altre valor de descriptor únic millor que el que ara tenim.
- Dotar l'aplicació de la possibilitat de poder treballar amb imatges que continguin múltiples figures. En l'àmbit mèdic podria ser interessant per fer classificació de microcalcificacions.
- Permetre que no sempre calgui tenir el mateix nombre de descriptors de Fourier per comparar dues imatges. Es podria fer que mentre la imatge estigués correctament mostrejada, quan es volguessin comparar imatges amb diferent nombre de descriptors de Fourier es busquessin les mostres que falten amb alguna tècnica d'interpolació.
- Mètode d'aprenentatge automàtic, imposant un llindar a partir del qual es pot considerar que una imatge no s'assembla a res i per tant és nova i si es volgués que es pogués afegir a la base de dades per reconèixer-la en cas que es preguntés per imatges semblants.