

INFN-MED

MAGIC-5

Publicazione su WWW del DB Mammografico e Polmonare

1. Introduzione

L'attività di ricerca nel campo dell'analisi di immagini medicali attualmente portata avanti nel contesto di MAGIC-5 e' stata iniziata nel 1998 dal progetto CALMA. Nel corso degli anni ha condotto allo sviluppo di prototipi di CAD mammografico e polmonare, attualmente in fase di validazione e/o perfezionamento.

Tutto lo sviluppo degli algoritmi si basa sulla disponibilita' di database di immagini digitali(zzate) a cui sia associato il cosiddetto "ground truth", che nel caso delle mammografie e' costituito dal risultato della biopsia per i positivi e dal *follow-up* per i negativi. nel caso delle CT polmonari invece, vista l'assenza di criteri oggettivi derivanti da ulteriori approfondimenti diagnostici (quali ad esempio quelli biotipici, istologici o citologici), che in questo particolare percorso diagnostico solitamente non vengono effettuati prima che l'eventuale nodulo sospetto abbia raggiunto dimensioni ragguardevoli, i criteri saranno concordati con i radiologi in conformita' con i piu' accreditati protocolli internazionali.

Per questo motivo, sia per quanto riguarda le mammografie che le CT polmonari, sono stati creati dei DB di ragguardevoli dimensioni (circa 4000 mammografie, circa 160 scansioni da CT polmonare per un totale di circa 48000 immagini di singole *slices*), utilizzati per lo sviluppo, il test e la validazione degli algoritmi. La loro qualita' e le loro dimensioni sono tali da poterne fare dei punti di riferimento utilizzati dalla comunita' internazionale.

A questa ragione, di per se' sufficiente a giustificarne la pubblicazione su Internet, si aggiunge il fatto che la validazione su un DB di accesso pubblico sta diventando un requisito necessario per la pubblicazione su rivista dei risultati, come verificato recentemente dai referaggi ricevuti per i nostri lavori da parte di riviste del settore di ampia rilevanza internazionale (quale ad esempio Medical Physics).

2. Mammografia

Nell'ambito dell'analisi di immagini mammografiche, riteniamo che sia importante perseguire l'obiettivo di rendere disponibili gli strumenti sviluppati agli utenti interessati.

Per quanto riguarda il database, in particolare, oltre alle immagini digitalizzate provenienti dal progetto CALMA, che sono state rivedute e tradotte in formato DICOM (lo standard in medicina), disponiamo di circa 300 immagini digitali fornite dal programma di *screening* della regione Umbria. Inoltre, grazie alla collaborazione con l'Ospedale Valdese di Torino, oggi dotato di 2 mammografi digitali, abbiamo accesso a circa 4000 immagini/mese per eventuali ampliamenti.

Lo sviluppo di CAD mammografici, in questa fase, si propone un obiettivo fondamentale, essenziale per qualunque futura applicazione, soprattutto su dati distribuiti: dimostrare che la risposta dell'algoritmo da validare sia omogenea rispetto a immagini provenienti da mammografi diversi, quindi con caratteristiche diverse (*range* dinamico, *pixel pitch*, etc.).

Riteniamo che mettere a disposizione della comunità internazionale il DB INFN sia importante per favorire lo sviluppo di migliori strumenti CAD e per contribuire a finalizzare una attività che è stata impegnativa sia in fase di raccolta che di refertazione e revisione.

Inoltre, a seguito della pubblicazione del lavoro descrittivo del DB, abbiamo ricevuto alcune richieste di accesso ai dati, che ad oggi non siamo stati in grado di soddisfare.

3. CT polmonari

Un database completo ed affidabile di CT del torace costituisce uno strumento di lavoro indispensabile sia per la realizzazione e l'ottimizzazione di algoritmi di analisi automatica di tali immagini sia per il loro test. La particolare necessità di poter disporre di un tale strumento da parte degli sviluppatori di SW e di sistemi CAD nel campo della diagnosi (in particolar modo precoce) delle neoplasie polmonari è giustificata dalle peculiarità proprie di questo ambito di ricerca [referenza al proposal INFN?].

In particolare, le peculiari caratteristiche del percorso diagnostico in oggetto rendono particolarmente critica la fase di determinazione radiologica della presenza di noduli sospetti e la valutazione del loro incremento dimensionale nelle CT di *follow-up*, e rendono quindi assai delicata la fase di annotazione e refertazione delle scansioni che possano costituire un database di riferimento. Infatti nella stragrande maggioranza dei casi non si ha a disposizione una "verità" biotipica, citologica o istologica. È per questo motivo che nella comunità radiologica internazionale è ancora aperto e di grande attualità anche il dibattito relativo alla determinazione di un "gold standard" per i database di riferimento e dei percorsi di *Quality Assurance* da intraprendere per garantire la massima affidabilità ai database in oggetto [S.G. Armato et al., Acad. Radiol., vol.14, pagg. 1455-1463, 2007].

A livello internazionale, attualmente gli unici database pubblici di CT polmonari sono quello reso disponibile dal consorzio LIDC [S.F. McNitt-Gray et al., Acad. Radiol., vol.14, pagg. 1464-1474, 2007] e quello proposto per la competizione ANODE09 [<http://anode09.isi.uu.nl/>].

Il Lung Image database Consortium (LIDC) è un progetto che coinvolge quattro tra le più importanti università americane, sotto l'egida del National Cancer Institute (NCI) con lo scopo di sviluppare delle linee guida nella raccolta, nella valutazione di CT polmonari e nella costruzione di un database di immagini relative, nella diagnosi del cancro al polmone. Esso mette a disposizione, gratuitamente, immagini e *trial* per la valutazione di sistema CAD e per l'addestramento di giovani radiologi. Le CT a disposizione sono ora 85, disomogenee sia per parametri di acquisizione che di ricostruzione, e sono accompagnate dalle refertazioni di 4 radiologi senza consensus.

La competizione ANODE09 consiste nella proposta di validare sistemi CAD su un database di 50 lung CT proveniente dal trial olandese "Nelson study" [<http://www.nelsonproject.nl/index.asp?index=index>] omogenee sia per parametri di acquisizione che di ricostruzione, di cui è ignota la refertazione. Sono inoltre disponibili 5 CT annotate di prova.

Per quanto riguarda il panorama italiano, nell'ambito della collaborazione MAGIC-5, la raccolta e refertazione di CT polmonari, sia provenienti dal progetto di *screening* ITALUNG_CT che cliniche ha costituito una delle attività di ricerca degli scorsi anni. Attualmente il database disponibile consta in totale di più di 160 CT a bassa dose ed alta risoluzione, omogenee sia per parametri di acquisizione che di ricostruzione, fra le quali sono state selezionate 50 CT provenienti dal progetto ITALUNG_CT che attualmente costituiscono il "*database gold standard*" che costituisce lo strumento mediante il quale vengono effettuati i confronti delle prestazioni dei vari algoritmi di analisi sviluppati nell'ambito della collaborazione. La realizzazione di questo database è stata resa possibile dalla preventiva realizzazione ed implementazione di una stazione dedicata di visualizzazione e refertazione delle CT polmonari (LUNG Nodule Annotation tool, detto LUNA), ed ha richiesto per l'annotazione completa con la nostra stazione dedicata un lavoro extra (cioè escludendo i tempi di acquisizione e ricostruzione delle immagini) da parte dei radiologi di più di 200 ore.

È intenzione della collaborazione proseguire nel popolamento del database sia di CT fornite da ITALUNG_CT (in particolare proseguirà anche l'acquisizione di esami di *follow-up*, necessari anche per lo sviluppo dei *tools* di analisi dimensionale) che cliniche, nella prospettiva (le cui modalità devono essere concordate con i radiologi), di realizzare un database pubblico di CT annotate che possa essere reso disponibile per la comunità scientifica internazionale. A tale scopo, oltre all'acquisizione delle immagini ed alla loro annotazione radiologica, sarà necessario un continuo aggiornamento, da effettuarsi in stretta collaborazione con i radiologi stessi, al fine di stabilire ed applicare al meglio protocollo di annotazione e criteri di determinazione del *gold standard* (numero di refertatori e loro caratteristiche ed esperienza, eventuale consensus, lettura blinded o no, classi di dimensioni, categoria e rilevanza dei noduli, criteri di *Quality Assurance*).

4. Programma di lavoro

L'obiettivo di questa proposta è di realizzare un portale WEB, accessibile agli utenti registrati, su cui pubblicare le immagini del DB mammografico e polmonare dell'INFN e di predisporlo per una futura gestione.

Per quanto riguarda il DB mammografico occorre:

- rivedere la correttezza delle diagnosi e della loro posizione in collaborazione con i radiologi;
- predisporre la struttura generale del portale;
- implementare i moduli per la registrazione/autenticazione degli utenti;
- implementare uno script per la certificazione e pubblicazione sul DB degli esami;
- implementare l'interfaccia dello script con la stazione GPCALMA, in modo tale da automatizzare la pubblicazione di nuovi casi, di cui ora non è disponibile la refertazione.

Per quanto riguarda il DB polmonare, è necessario realizzare funzionalità analoghe, con una implementazione separata in quanto ci si rivolge ad una diversa comunità di utenti.

Allo scopo di raggiungere gli obiettivi sintetizzati in precedenza, è necessario l'impegno di alcuni componenti della collaborazione MAGIC-5, dei radiologi e di una persona full-time per circa 2 anni (uno per la implementazione ed il test, uno per l'inizio della gestione ed il training dei componenti della collaborazione che gestiranno l'operatività del portale).