TESI DOCTORAL

Títol	Ajuda al Diagnòstic de Càncer de Melanoma amb Raonament Analògic Multietiqueta
Realitzada per	Rubén Nicolàs i Sans
en el Centre	Escola Tècnica Superior d'Enginyeria Electrònica i Informàtica La Salle
i en el Departament	Informàtica
Dirigida per	Dra. Elisabet Golobardes i Ribé Dr. Albert Fornells Herrera

C. Claravall, 1-3
08022 Barcelona

Ajuda al Diagnòstic de Càncer de Melanoma amb Raonament Analògic Multietiqueta

Per

Rubén Nicolàs i Sans

Dirigida per

Dra. Elisabet Golobardes i Ribé
Dr. Albert Fornells Herrera

Una Tesi Sotmesa a
ETSEEI La Salle - Universitat Ramon Llull
per al grau de
Philosophiae Doctor (Ph.D.)
Programa de Doctorat
Tecnologies de la Informació i les Comunicacions i la seva Gestió
Febrer 2014

A Carles, Esther, Pili i Alba.

"Si vis amari ama"

Aurelius Agustinus

Agraïments

Una tesi doctoral és un document molt personal. És difícil d'escriure ja que ha d'abraçar la feina i els esforços de molts anys. Tot i que les característiques personals i de dificultat d'escriptura són globals, s'acusen especialment en aquest capítol. Si ja costa resumir els anys de feina, és gairebé impossible resumir els anys viscuts. Son moltes les cares que passen pel cap en pensar amb qui s'ha fet camí. Hi són presents en aquest repàs mental la família, els companys, els amics, els referents i tots aquells que amb la seva presència, la seva absència, la seva paraula o el seu soroll de fons han aportat llum a aquest treball.

Vull començar essent agraït amb la família, per fer-se present quan l'he requerit i, a més, saber no ser-hi quan necessitava solitud. El meu més profund agraïment als pares, als avis i a qui m'ha acompanyat els darrers anys.

Per continuar vull donar les gràcies a tots els companys que han col·laborat amb mi mentre em barallava per donar a llum aquest treball. Als que m'han guiat, als que m'han aportat idees i als pocs que seguint aquest sender han sabut passar de col·legues a amics.

Vull agrair l'ajuda dels que han fet confiança a la present tesi: A. Càrceles, C. Nicolàs, E. Sans, E. Jover, E. Golobardes, A. Fornells, JM. Campora, FX. Babot, G. Inglés, S. Mur, M. Pellicer, N. Macià, L. Kinnear, L. Moya, J. Rios ... Sense ells no hagués pogut saltar els obstacles d'aquesta cursa.

Per últim, a tots els que us hi sentiu representats.

Gràcies.

Resum

La mortalitat provocada pel càncer de melanoma ha augmentat en els últims anys a causa, principalment, dels nous hàbits d'exposició al sol. Atenent al criteri mèdic, el diagnòstic precoç s'ha convertit en el millor mètode de prevenció. No és però una tasca trivial ja que els experts del domini han de fer front a un problema caracteritzat per tenir un gran volum de dades, de format heterogeni i amb coneixement parcial. A partir d'aquestes necessitats es proposa la creació d'una eina de suport a la presa de decisions que sigui capaç d'ajudar els experts en melanoma en el seu diagnòstic. El sistema ha de fer front a diversos reptes plantejats, que inclouen la caracterització del domini, la identificació de patrons a les dades segons el criteri dels experts, la classificació de nous pacients i la capacitat d'explicar els pronòstics obtinguts. Aquestes fites s'han materialitzat en la plataforma DERMA, la qual està basada en la col·laboració de diversos subsistemes de raonament analògic multietiqueta. L'experimentació realitzada amb el sistema proposat utilitzant dades d'imatges confocals i dermatoscòpiques ha permès comprovar la fiabilitat del sistema. Els resultats obtinguts han estat validats pels experts en el diagnòstic del melanoma considerant-los positius.

Paraules clau. Ajuda al diagnòstic, càncer de melanoma, raonament analògic, sistemes col·laboratius, classificació multietiqueta.

Resumen

La mortalidad a causa del cáncer de melanoma ha aumentado en los últimos años debido, principalmente, a los nuevos hábitos de exposición al sol. Atendiendo al criterio médico, el diagnóstico precoz se ha convertido en el mejor método de prevención, pero no se trata de una tarea trivial puesto que los expertos del dominio deben hacer frente a un problema caracterizado por tener un gran volumen de datos, de formato heterogéneo y con conocimiento parcial. A partir de estas necesidades se propone la creación de una herramienta de ayuda a la toma de decisiones que sea capaz de ayudar a los expertos en melanoma en su diagnóstico. El sistema tiene que hacer frente a diversos retos planteados, que incluyen la caracterización del dominio, la identificación de patrones en los datos según el criterio médico, la clasificación de nuevos pacientes y la capacidad de explicar los pronósticos obtenidos. Estas metas se han materializado en la plataforma DERMA la cual está basada en la colaboración de varios subsistemas de razonamiento analógico multietiqueta. La experimentación realizada con el sistema propuesto utilizando datos de imágenes confocales y dermatoscópicas ha permitido verificar la fiabilidad del sistema. Los resultados obtenidos han sido validados por los expertos en el diagnóstico del melanoma considerándolos positivos.

Palabras clave. Ayuda al diagnóstico, cáncer de melanoma, razonamiento analógico, sistemas colaborativos, clasificación multietiqueta.

Abstract

Mortality related to melanoma cancer has increased in recent years, mainly due to new habits of sun exposure. Considering the medical criteria, early diagnosis has become the best method of prevention but this is not trivial because experts are facing a problem characterized by a large volume of data, heterogeneous, and with partial knowledge. Based on these requirements we propose the creation of a decision support system that is able to assist experts in melanoma diagnosis. The system has to cope with various challenges, that include the characterization of the domain, the identification of data patterns attending to medical criteria, the classification of new patients, and the ability to explain predictions. These goals have been materialized in DERMA platform that is based on the collaboration of several analogical reasoning multi-label subsystems. The experiments conducted with the proposed system using confocal and dermoscopic images data have been allowed to ascertain the reliability of the system. The results have been validated by experts in diagnosis of melanoma considering it as positive.

Keywords. Diagnostic aid, melanoma cancer, analog reasoning, collaborative systems, multi-label classification.

Tesi doctoral per compendi de publicacions

La present tesi doctoral s'acull a la normativa per a l'elaboració de tesis doctorals per compendi de publicacions de la Universitat Ramon Llull[1]. La normativa consta dels següents punts:

1. Una tesi doctoral per compendi de publicacions estarà formada per un mínim de tres articles sobre una mateixa línia d'investigació.

2. Només s'acceptaran articles de publicacions que disposin d'un sistema d'avaluació per *peer review* i/o que estiguin indexades preferentment en bases de dades científiques internacionals.

3. Només s'acceptaran articles publicats, o acceptats per a la seva publicació, realitzats amb data posterior a la primera matriculació del doctorand als estudis de doctorat o màster oficial.

4. Els coautors dels articles publicats donaran la seva conformitat per escrit a la utilització de l'article com a part de la tesi del doctorand.

5. Els coautors dels articles publicats no formaran part del tribunal de la tesi.

6. Els coautors dels articles publicats i utilitzats en una tesi que no tinguin el grau de doctor renunciaran per escrit a utilitzar l'article en una altra tesi. En el cas que els articles publicats siguin de més d'un equip de recerca, la Comissió de Doctorat del centre podrà considerar excepcions justificades en l'aplicació d'aquesta norma.

7. La tesi comptarà amb una introducció general que presenti els treballs publicats, una justificació de la unitat temàtica, una còpia de cada treball publicat, un resum global dels resultats, la seva discussió i les conclusions finals.

8. Per tot el citat anteriorment, s'haurà de presentar sempre, a l'inici del procés de la tesi, una sol·licitud formal a la Comissió de Doctorat del centre i obtenir la seva acceptació favorable. La Comissió vetllarà per la qualitat de les publicacions que es volen presentar per a la Tesi. A la sol·licitud s'afegirà també un informe del director de la Tesi indicant quina és la contribució específica del doctorand al treball presentat i la de la resta d'autors, si s'escau. S'haurà de presentar l'acta d'aprovació de la Comissió del centre a la Comissió de Doctorat de la URL en el moment de la tramitació ordinària de la Tesi.

[1] Aprovada per la Junta Acadèmica a 18 de setembre de 2008

Aquesta tesi compleix amb tots els punts prèviament citats. Les tres publicacions que formen el compendi són les següents:

1. R. Nicolas, A. Fornells, E. Golobardes, G. Corral, S. Puig, and J. Malvehy, DERMA: A melanoma diagnosis platform based on collaborative multi-label analog reasoning. *The Scientific World Journal*, 2014. Publicació indexada amb factor d'impacte 1,730.

2. R. Nicolas, A. Sancho-Asensio, E. Golobardes, A. Fornells, and A. Orriols-Puig, Multi-label classification based on analog reasoning. *Expert Systems With Applications Journal*, 2013. Publicació indexada amb factor d'impacte 1,854.

3. R. Nicolas, D. Vernet, E. Golobardes, A. Fornells, S. Puig, and J. Malvehy, Improving the combination of CBR systems with preprocessing rules in melanoma domain. In *8th International Conference on Case-Based Reasoning, Workshop on Case-Based Reasoning in the Health Sciences*, 2009. Publicació Internacional Core C.

Índex

Índex de figures

Índex de taules

"No hi ha cap vent favorable per aquell que no sap a quin port es dirigeix"

Arthur Schopenhauer

1

Marc de treball, motivació i objectius

El càncer de melanoma és una malaltia que afecta cada cop més persones, per tant, tot allò que pugui aportar millores en el seu diagnòstic precoç esdevé de vital importància, més encara considerant que l'eficiència del tractament augmenta en les fases inicials de la malaltia. Aquesta tesi tracta sobre el plantejament de DERMA, una eina d'ajuda a la presa de decisions en el diagnòstic de melanoma. En aquest capítol s'analitzarà què ha motivat el treball i el context en què s'ha treballat. Veurem com la participació en diferents projectes ha permès la col·laboració amb els experts mèdics i ha marcat els objectius principals de la recerca.

1.1 Marc de treball

La present tesi s'inicia l'any 2006 amb una beca de recerca concedida per La Salle, Universitat Ramon Llull. La beca es va substituir, l'any 2007, per l'ajut de Formació d'Investigadors (2007FI_A 01328), concedit per l'Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR) de la Generalitat de Catalunya, i cofinançat pel Fons Social Europeu. La recerca s'ha realitzat en el si del Grup de Recerca en Sistemes Intel·ligents (GRSI) de La Salle, un grup de recerca consolidat, reconegut i finançat per la Generalitat de Catalunya des de l'any 2002 (2002-SGR-00155, 2005-SGR-00302 i 2009-SGR-183). La investigació del GRSI se centra en la Intel·ligència Artificial / *Artificial Intelligence* (AI) (Russell et al., 1996), sobretot en el camp de la Mineria de Dades / *Data Mining* (DM) (Maimon and Rokach, 2005), i és reconegut per la seva experiència en Raonament Basat en Casos / *Case-Based Reasoning* (CBR) (Aamodt and Plaza, 1994), Computació Evolutiva / *Evolutionary Computation* (EC) (Jong, 2006) i

Xarxes Neuronals / *Neural Networks* (NN) (McCulloch and Pitts, 1988). Algunes de les aplicacions recents del GRSI són l'anàlisi assistida d'atacs en xarxes informàtiques, l'educació, l'energia i, sobretot, les eines de suport al diagnòstic mèdic.

Com veurem més endavant, la tesi s'emmarca en el domini mèdic. L'aplicació de tècniques d'AI a problemes mèdics es deu al fet que aquest tipus de problemes tenen volums d'informació que no són directament processables pels mètodes convencionals. La dificultat pot ser coberta per tècniques que facin que una màquina tingui un comportament que seria considerat intel·ligent en un humà, tret fonamental de l'AI. Aquest és, a més, un àmbit d'investigació ben reconegut, amb el suport de diferents institucions i certàmens com la *Google Science Fair 2012*, que va premiar la recerca del càncer de mama, o l'*Intel International Science and Engineering Fair 2013*, que va guanyar un projecte sobre leucèmia.

Des de l'òptica del GRSI, i dels projectes en els quals col·labora, es treballa en els dominis mèdics a través de DM. La família de DM té el focus en l'extracció de coneixement processable implícit en les dades. La gamma de tècniques de mineria de dades aplicades comprenen quatre grups principals de mètodes: clusterització, regles d'associació, classificació i regressió. En quant al grup de la clusterització, l'objectiu és descompondre el problema i tractar de modelar-lo d'una manera adequada. La clusterització en l'àmbit mèdic pot aplicar-se a l'agrupació de pacients amb símptomes similars on cada grup caracteritza un estadi de la malaltia. En el cas de les regles d'associació, l'interès se centra en els conjunts de característiques que en fan succeir altres. Les regles poden definir quins atributs d'un pacient comporten un determinat tipus de malaltia complementant els patrons mèdics. La classificació permet donar una tipologia de problema a través de les característiques d'un nou cas, fet que té grans similituds amb el procés de diagnòstic mèdic. A través de la classificació es pot definir el tipus d'un nou pacient a partir de les seves dades. La regressió, per últim, mostra la forma com les variables regressores expliquen la variable resposta i fa pronòstics dels valors d'aquesta variable permetent, en els dominis mèdics, fer prediccions de la malaltia.

La recerca realitzada en la tesi no s'emmarca només en el GRSI sinó en un dels seus projectes, concretament, en la participació entre el 2006 i el 2009 al projecte *Marco Integrador para el Desarrollo de Sistemas de CBR* (MID-CBR)(TIN2006-15140-C03), finançat pel *Ministerio de Ciencia y Tecnología* i *Fondo Europeo de Desarrollo Regional*. El focus del projecte està en el CBR, una tècnica per resoldre nous problemes utilitzant-ne altres de prèviament resolts. A partir de les característiques d'un nou problema s'analitzen les situacions prèviament experimentades tot cercant les que siguin més similars per establir solucions anàlogues que resolguin la problemàtica proposada. Per assolir els objectius de funcionament s'aplica un cicle de quatre fases (recuperació, adaptació, revisió i emmagatzematge) que es nodreixen de la Memòria de Casos / *Case Memory* (CM), on hi ha les experiències anteriors de sistema. El projecte es va dur a terme a partir de la col·laboració entre l'*IIIA - Consejo Superior de Investigaciones Científicas*, *GAIA - Universidad Complutense de Madrid* i el GRSI. Els objectius del projecte al GRSI se centren a abordar problemes, mitjançant CBR, principalment de classificació, que siguin complexos, amb coneixement imprecís o amb gran volum de

dades. Concretament, emprendrem el tractament de la recuperació per a CBR amb ús intensiu de dades. Aquests tipus de problemes són interessants per la seva dificultat de tractament amb els mètodes tradicionals. En aquesta tasca del projecte s'aborden dues línies d'investigació per millorar la precisió en la recuperació de casos: l'EC i l'Aprenentatge Conjunt / *Ensemble Learning* (EL) (Dietterich, 2002). Les dues tècniques tenen en comú l'exploració paral·lela de diferents possibilitats per trobar-ne l'òptima, malgrat que es realitza de manera diferent. La computació evolutiva realitza una recerca paral·lela simultània en espais de cerca diferents, mentre que l'EL parteix les dades en bases de casos paral·leles i, posteriorment, n'agrega els resultats. L'equip del GRSI aporta dos dominis d'experimentació al projecte MID-CBR: un en l'àmbit de la medicina, i més concretament en el del càncer de melanoma, i un altre en el de la telemàtica. Per les característiques del grup de treball, el domini d'aplicació assignat va ser el mèdic i es va iniciar la col·laboració amb la Unitat de Melanoma de l'Hospital Clínic i Provincial de Barcelona (HCPB) - Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Ens Promotor Observador (EPO) del projecte.

Addicionalment es va considerar interessant ampliar el coneixement en àrees concretes atenent a les necessitats del problema. D'aquesta manera, l'any 2008 es va participar en el segon *International Summer Course on Soft Computing: Intelligent Data Analysis*, de l'*European Centre for Soft-Computing*, Mieres, Astúries. El curs va permetre aprofundir en els coneixements de l'anàlisi de dades per tal d'aplicar-los al domini del melanoma. En concret, es van estudiar tècniques per a la caracterització del domini. A més, l'any 2009, i durant set mesos, es va realitzar una estada al *Component Analysis and Human Sensing Lab*, de *Carnegie Mellon University* (CMU) a Pittsburgh, PA, USA, sota la supervisió del cap del grup, Fernando de la Torre. Durant aquest període es va seguir aprofundint en l'estudi de mètodes d'anàlisi de dades per aplicar-los a la millora de les memòries de casos. Les dues formacions van permetre una major especialització en els mètodes a aplicar en el domini del melanoma.

1.2 Motivació

El càncer de melanoma ha crescut en els darrers anys i és una malaltia cada cop més freqüent en la nostra societat que afecta persones de qualsevol edat. Encara que el melanoma no és el càncer de pell més comú, la seva mortalitat és elevada. Segons l'Acadèmia Americana de Dermatologia / *American Academy Dermatology* (AAD) (McWhirter and Hoffman-Goetz, 2013) causa gairebé un setanta-cinc per cent de les morts per càncer de pell, esdevenint encara més mortal si no es tracta en les seves primeres etapes. Com en la majoria de càncers, la millor estratègia per abordar-lo és el diagnòstic precoç. Ara bé, la dificultat en el diagnòstic precoç recau en el gran volum d'informació a considerar, els diferents tipus d'experts implicats (entre altres els dermatòlegs, els oncòlegs i els patòlegs) i diagnòstics implicats i el fet que no es disposa de protocols complets adients per a tots els actors del procés. Aquest conjunt de característiques ha induït a l'aplicació de tècniques de DM per explotar les dades i ajudar els experts en el seu diagnòstic. Tenint en compte la bibliografia de DM en

medicina podem diferenciar dos tipus d'enfocament a l'hora d'aplicar les tècniques: el dels mètodes per trobar patrons i el de les eines per ajudar a la presa de decisions. A continuació, veurem alguns dels principals exemples de cada grup en l'àmbit mèdic.

Les tècniques de DM s'utilitzen per trobar patrons que permetin crear eines de suport a la decisió. La finalitat d'ajudar a la presa de decisions s'ha tractat, en l'àmbit del càncer, amb diferents tècniques i des de diversos punts de vista com exposa Jain et al. (2000). El primer objectiu mèdic és aconseguir una extracció de característiques automàtica per detallar les propietats dels problemes. La caracterització sol fer-se a través d'un procés de clusterització (Han and Kamber, 2006) com es detalla en el treball realitzat per Singh et al. (2011). A més, per evitar tècniques invasives que tenen conseqüències negatives en els àmbits psicològic, de la salut i econòmic, es necessita ampliar l'ús de tècniques com la termografia o l'anàlisi d'imatges per a l'ajuda al diagnòstic. En aquesta línia, Cruz-Ramírez et al. (2013) presenten un enfocament que utilitza xarxes baiesianes (Jensen and Nielsen, 2007) per modelitzar la informació termogràfica i poder així establir classificacions a través de processos no invasius. S'ha de destacar que els processos de modelat de la informació han de prioritzar la fiabilitat. Un exemple de la voluntat de precisió és la plataforma DESMAI, proposada per Fornells et al. (2008b), que utilitza la clusterització per organitzar la informació i permetre recuperar-la d'una manera més ràpida i fiable. En el cas concret del càncer de melanoma, un dels mètodes més destacats consisteix a construir un model del domini combinant mètodes d'aprenentatge de diferents característiques (Armengol and Puig, 2011). En l'àmbit de la recomanació en melanoma i, més concretament, en l'àmbit del descobriment de coneixement, destaca el treball de Fornells et al. (2008a), que utilitza la clusterització per a identificar grups similars de melanomes i crear descripcions dels grups que s'utilitzaran com a explicacions per als experts. Armengol (2011) utilitza una combinació de CBR i clusterització per a la generació d'una teoria del domini per classificar els melanomes in situ. Aquest procés es realitza per donar suport als dermatòlegs en l'avaluació de lesions de la pell utilitzant dermatoscòpia i abans de l'extracció.

Amb aquests precedents, els principals reptes a afrontar en la creació d'una eina d'ajuda al diagnòstic de melanoma són: l'extracció de característiques del domini, l'organització de la informació, la classificació amb ús de tècniques no invasives, la fiabilitat i l'explicació dels resultats obtinguts. La motivació de la tesi se centra en aquests punts que, com hem anat veient, han estat abordats de manera individual per alguns dels sistemes anteriorment explicats. Amb tot, les propostes dels mètodes presentats no aporten la visió de conjunt que demanen els experts. A més, els especialistes volen un sistema que tingui com a motor de classificació el seu protocol de decisió, fet que no correspon exactament amb cap de les tècniques disponibles. Per tant, no es poden utilitzar els mètodes plantejats fins ara o, si més no, no directament. Per tot això s'ha de dissenyar una proposta que combini diferents tècniques de DM amb el protocol mèdic de diagnòstic en melanoma.

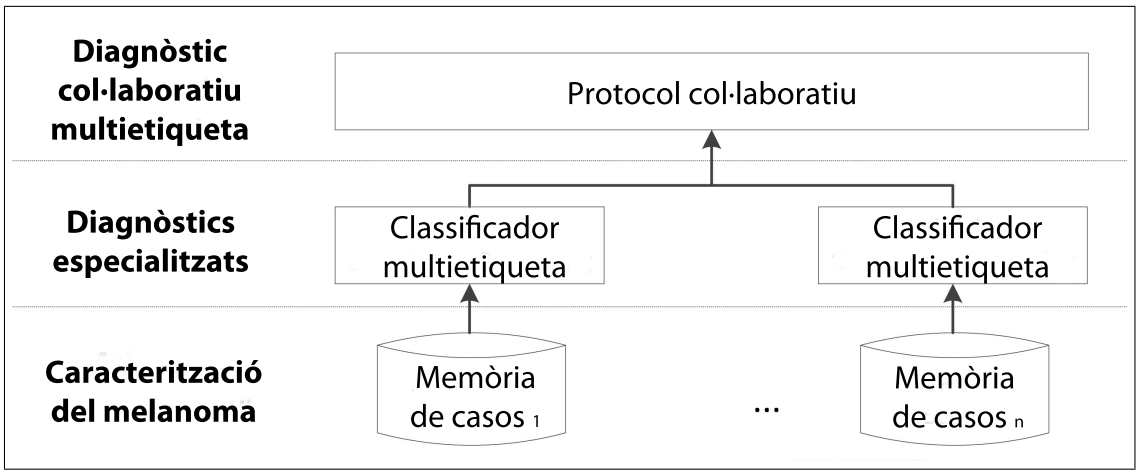

Figura 1.1: Blocs funcionals de l'eina d'ajuda al diagnòstic. Mostra les tres capes del model funcional i els diferents mòduls de cada capa.

1.3 Objectius

Amb les motivacions com a horitzó observem que la fita global és dissenyar un sistema capaç d'ajudar els experts en càncer de melanoma en el seu procés de diagnòstic. La figura 1.1 mostra l'arquitectura del sistema que es proposa, *Melanoma Diagnosis based on Collaborative Multi-Label Reasoning* (DERMA), el qual ha de fer front a quatre fases que són, també, les quatre aportacions principals de la tesi:

Caracterització del domini. El primer objectiu és establir quines són les característiques del domini atès que actualment cada expert utilitza les seves fonts de dades etiquetades segons la seva terminologia. La manca d'una estructura global de dades propicia la duplicitat d'informació i la dificultat de col·laboració diagnòstica entre els diferents professionals implicats. Com es descriu a Nicolas et al. (2008a) existeixen moltes propostes de bases de dades mèdiques, fins i tot dins l'àmbit del càncer de melanoma, però cap d'elles dóna una visió global de les dades i com es relacionen. És per això que, conjuntament amb la Xarxa Catalana de Melanoma (XCM), es dissenya una base de dades que inclogui totes les dades i experts de càncer de melanoma. La caracterització del domini s'estableix a través d'entrevistes amb experts i de l'estudi dels patrons de la malaltia. Amb aquest objectiu s'aconsegueix un model de dades complet on es detalla tota la informació disponible i les seves relacions. El model ha de permetre la col·laboració entre experts i evitar la pèrdua i duplicació d'informació. Com es veu a la figura 1.1 l'etapa de caracterització serà la base de l'eina d'ajuda al diagnòstic ja que les dades aconseguides conformen les CM del sistema.

Diagnòstics parcials no col·laboratius. Un cop definit el domini, s'ha de determinar per a cada expert i tipus d'informació quina és la classificació més adient optimitzant els conjunts de dades i la seva recuperació. Es fa front, per tant, a

un problema de classificació amb un tret diferencial respecte a altres propostes de DM, com és l'ús del protocol mèdic com a motor de classificació. D'aquesta manera es pretén que la classificació de nous pacients pugui fer-se seguint estrictament l'algorisme utilitzat pels experts però afegint-hi la potencialitat de l'ús de la computació. El segon repte és, per tant, la creació de subsistemes especialitzats per treballar amb les diferents fonts de dades. El fruit d'aquest punt ha de ser una eina que permeti fer una classificació de pacients atenent a un expert o prova concreta. La classificació ha de ser el més acurada possible, per això en aquest pas també es considera la forma d'organitzar les bases de coneixement d'una manera apropiada per millorar-ne el rendiment. A la figura 1.1 hi veiem com cadascun dels diagnòstics especialitzats se sustenten d'una CM. Aquesta característica permet que cada expert classifiqui només en el seu tipus de diagnòstic i amb la seva informació, una especialització que també es produeix en el procediment mèdic.

Diagnòstic global col·laboratiu. De l'estudi del problema en destaca que no només interessa fer l'anàlisi correcta de les dades i la seva classificació sinó que es busca, per sobre de tot, la fiabilitat. I en el procediment mèdic la fiabilitat s'obté a través de la unió de les diferents tècniques i especialistes. D'aquesta manera, cal establir la classificació a partir de la col·laboració entre els experts i les proves. Com en el punt anterior, els experts demanden l'estricte seguiment del seu protocol de decisió i per això no podem utilitzar el model general de col·laboració (Melville and Mooney, 2004). La seva petició motiva una adaptació dels mètodes actuals al protocol mèdic. El tercer repte és, doncs, definir un esquema de col·laboració entre els subsistemes de classificació independents fent servir com a base la manera en què els experts treballen i que també inclou mecanismes per a la gestió de situacions excepcionals. El sistema resultant ha de permetre donar una classificació d'un pacient coordinant les opinions dels diferents experts i proves. La capa col·laborativa de l'eina, com es veu a la figura 1.1, actua unificant els criteris de les classificacions prèvies i, com en les altres etapes, es fonamenta en un procediment mèdic: les reunions de consens entre experts per analitzar casos concrets.

Diagnòstic multietiqueta. La classificació del càncer de melanoma ha de ser eficient, considerant tots els agents implicats, i amb la possibilitat d'aportar resultats amb més d'un tipus d'informació. Les diferents propostes en mètodes de classificació multietiqueta permeten una bona classificació. Tot i així, fins a la data no s'han proposat treballs que adaptin algorismes d'aprenentatge analògic que es puguin utilitzar en el problema del melanoma a la classificació de més d'una etiqueta (Nicolas et al., 2013b, 2014). El treball més similar, Zhang and Zhou (2005), proposa una versió per a múltiples etiquetes de K-Veí més Proper / *K-Nearest Neighbor* (KNN) (Han and Kamber, 2006) però no s'adapta a les característiques concretes del problema estudiat doncs es desvia del protocol de diagnòstic mèdic. Així, s'haurà d'adaptar el mètode específic obtingut dels anteriors reptes a aquesta nova necessitat. La fita, com s'aprecia a la figura 1.1, resta

a cavall entre el diagnòstic especialitzat, que ha de passar a ser multietiqueta, i el diagnòstic col·laboratiu que ha de fer la unificació de múltiples etiquetes en lloc d'una.

1.4 Estructura de la tesi

La present tesi es divideix en tres blocs. El primer és el capítol introductori, en què es detalla el marc de treball, la motivació i els objectius de la tesi. El segon bloc és l'apartat d'aportacions, que conté quatre parts: el capítol 2 de caracterització del domini, el capítol 3 de diagnòstic no col·laboratiu, el capítol 4 de diagnòstic col·laboratiu i el capítol 5 amb la versió multietiqueta. Per últim, descrites les diferents fites, al capítol 6 s'hi fa l'anàlisi i discussió de resultats, les conclusions i les línies de futur. Finalment, els apèndixs contenen els acrònims utilitzats i les publicacions.

"Si no lluites tingues almenys la decència de respectar als qui sí que ho fan"

José Martí

2

Fita 1: Caracterització del domini

Un cop introduïda la problemàtica del diagnòstic del càncer de melanoma es descriuran les aportacions realitzades en aquesta tesi que es divideixen, com hem vist, en quatre fites. El present capítol aborda la primera de les fites que és la caracterització del domini. Aquest és el primer repte, puix que ens permetrà conèixer les característiques concretes del problema que volem tractar. Com veurem durant el capítol, a través de la col·laboració amb els experts de l'Hospital Clínic s'ha generat un model de les dades de càncer de melanoma amb la informació utilitzada pels experts i les seves interaccions. A més, s'explora la possibilitat d'una nova classificació del domini a través d'un procés de clusterització. A partir d'aquesta informació veurem, en els capítols posteriors, quins són els reptes per establir un mètode d'ajuda al diagnòstic a partir de les dades disponibles. Aquest procés, amb l'aplicació de tècniques d'intel·ligència artificial, és el que permetrà la creació d'una eina de classificació en melanoma, DERMA.

2.1 Motivació i objectius

El primer pas en el desenvolupament d'un sistema d'ajuda a la presa de decisions és identificar, entendre i reunir les dades associades al problema. Aquests passos no són trivials en un problema mèdic ja que les dades es caracteritzen per ser heterogènies i procedents de diversos perfils mèdics que es veuen implicats en el diagnòstic. A més, els experts etiqueten les dades d'acord a la nomenclatura pròpia de la seva especialitat, per tant, alguns atributs amb noms diferents poden tenir, realment, el mateix significat. Com a conseqüència, la caracterització i la comprensió de les relacions entre totes les

fonts de dades per a la planificació del coneixement és una tasca complexa que requereix consens entre els experts.

Amb l'objectiu d'assolir una caracterització exhaustiva del domini i les seves relacions es va analitzar tot el procés diagnòstic dels diferents investigadors de l'HCPB involucrats. Els experts afirmen que es tracta d'un domini amb gran disponibilitat d'informació, però que aquesta prové de fonts i formats heterogenis que en dificulten la integració. Fins ara, les dades estaven recollides en suports i en formats d'organització que no permetien una anàlisi conjunta. La manera de recollir i emmagatzemar les dades, així com la seva diversitat de tipus, va motivar la necessitat de crear una nova estructura de les dades que aglutinés de manera ordenada tot el coneixement del domini.

A més a més, a la mateixa caracterització de les dades, i per provar si era possible identificar patrons equivalents als utilitzats pels experts, es van aplicar tècniques de mineria de dades. Com veurem a la secció 2.3, el diagnòstic del melanoma considera les característiques que normalment s'observen en els nevus [1] per fer el diagnòstic. La tècnica, però, no s'ajusta exactament a tots els casos i, per això, els experts estan interessats a explorar noves classificacions.

A continuació veurem el detall de les dues aportacions: la caracterització del domini i la cerca de nous patrons de classificació.

2.2 Caracterització del domini

Fruit de la col·laboració amb l'HCPB es va aconseguir una definició del domini (Nicolas et al., 2008a). La caracterització del problema permet veure tota la informació utilitzada i com es relaciona i, a més, s'ha de ressaltar que es tracta d'un model únic ja que inclou dades de diversos anys d'experiència de l'HCPB i informació de tots els àmbits implicats en la malaltia. Fins aleshores només hi havia treballs amb dades parcials (clíniques o patològiques) o estudis amb informació particular d'alguns individus. És important destacar que si bé el diagnòstic precoç és de vital importància, també s'ha de tenir en consideració la seva precisió. Per poder assolir aquesta combinació de precisió i diagnòstic precoç, els experts necessiten un model relacional complet i consistent.

La caracterització del domini es va realitzar amb dades de més de tres mil pacients de l'HCPB incloses en els informes dels dermatòlegs, oncòlegs, cirurgians, patòlegs i altres especialistes que treballen en el centre. Les dades de les quals es disposa són heterogènies i estan distribuïdes en conjunts de dades independents, un per a cada expert. A més de trobar-se en bases de dades independents, bona part dels atributs són representats i emmagatzemats amb noms i formats diferents segons quin sigui l'expert que els utilitza. Això provoca que es tinguin dades iguals, o de la mateixa família, però amb noms diferents. Per aquesta raó, es va definir un model amb quaranta-sis conceptes, mitjançant el punt de vista i dades d'estudis internacionals que analitzen

[1] Proliferacions de cèl·lules pigmentades a la pell.

aspectes específics del domini i el divideixen en cinc grups (Nicolas et al., 2008a): (1) persona i família, (2) informació mèdica genèrica, (3) tumors, (4) metàstasis i (5) controls i estudis.

El modelat de dades i relacions va mostrar la complexitat de la informació i les restriccions del domini. Estudiant les dades, el model se centra en el pacient (filiació) a partir del qual s'obtenen diferents informacions com: (1) les relacions familiars amb altres pacients, per verificar el seu impacte en la malaltia; (2) els metges encarregats del tractament; (3) els tumors primaris amb les diferents intervencions i proves associades; (4) les metàstasis associades als tumors primaris i les accions preses respecte aquests tumors (relacionades o no amb el tumor primari); (5) el seguiment del pacient i els tumors; (6) les dades del pacient, des dels antecedents personals i els ingressos hospitalaris fins a les mostres emmagatzemades o el fototip i, finalment, (7) les informacions genètiques obtingudes del pacient. Aquesta caracterització detallada és la base de coneixement de la recerca dels experts en el domini.

Malgrat que l'estudi unificat de les dades permetia la integració de totes les fonts de coneixement, hi ha parts d'informació que no es podien utilitzar en una posterior fase d'experimentació, perquè els experts no tenien registres de tots els pacients. Així el model inclou tots els possibles aspectes a tractar en el diagnòstic de càncer de melanoma però no de tots ells es disposa de les dades d'un nombre ampli de pacients. Considerant aquesta manca d'uniformitat en les dades i la necessitat dels experts de testejar determinades tècniques diagnòstiques, es va fer un nou estudi del problema amb les dades utilitzades per a l'experimentació mèdica. Els conjunts de dades utilitzats es van centrar, llavors, en la informació pròpia dels nevus. Els nevus i, en concret, les seves imatges estan disponibles i centren l'anàlisi actual dels experts de l'HCPB. Més específicament, les imatges dermatoscòpiques i confocals van ser seleccionades per èsser les dues visualitzacions que ofereixen una millor descripció del melanoma. Com veurem en el capítol 6, aquest nou estudi de les dades va tenir com a resultat un conjunt de dades representatiu del domini amb 150 instàncies d'imatges dermatoscòpiques i confocals.

2.3 Extracció de patrons

Un cop modelitzat el domini, va sorgir un nou repte en la caracterització de les dades. El repte es fonamenta en la classificació actual dels nevus atenent a característiques concretes. Una de les tècniques diagnòstiques de melanoma més utilitzada actualment es basa en la regla Asimetria, Contorn, Color, Diàmetre / *Asymmetry, Border, Color, Diameter* (ABCD). Tot i que aquest criteri diagnostica correctament, hi ha un conjunt variable de casos que no acompleixen els criteris i, per aquest motiu, els experts en melanoma volen establir nous patrons que millorin la classificació. Amb tot, es busca que aquests patrons no tinguin un nombre prefixat de classes sinó que el nombre s'adapti per ser l'òptim.

Per arribar als objectius es planteja una proposta per identificar patrons útils per a la classificació de lesions en càncer de melanoma a partir d'un conjunt de dades creat

pels investigadors de l'HCPB. Els experts mèdics volen conèixer el nombre òptim de patrons de melanoma alhora que la seva caracterització. Per això s'ha triat el descobriment de patrons a partir de l'algorisme *k-means* (MacQueen, 1967) que permet fer una cerca incremental fent servir diferents nombres de clústers. Tant l'extracció com la validació d'aquests patrons s'ha fet a partir de l'eina *Unsupervised Learning In CBR* (ULIC) (Vernet and Golobardes, 2003) que fa una exploració intensiva de les dades a partir de tècniques de clusterització particional.

Com veurem en l'anàlisi de resultats (capítol 6), el procés d'extracció de patrons aconsegueix generar una nova manera de classificar els casos de melanoma. Atenent a criteris mèdics, els resultats van proporcionar pautes de classificació equivalents a les assenyalades per ABCD (Fornells et al., 2008a, Vernet et al., 2008).

2.4 Conclusions i passos següents

En aquest capítol hem vist quines són les característiques del domini del càncer de melanoma. Aquest modelat de la informació mèdica s'ha fet des de dos punts de vista. El primer segueix estrictament el model emprat pels experts en melanoma i s'ha fet a partir de la col·laboració amb l'HCPB. La caracterització té com a resultat un mapa del problema, que modelitza el conjunt de dades disponible i com les dades es relacionen entre elles. Del model, a posteriori, s'hi podrà extreure un sistema d'ajuda al diagnòstic. En segon lloc, s'han analitzat les dades disponibles amb un procés de clusterització per obtenir nous patrons de classificació sintètics que ajudin a afinar el diagnòstic. Els patrons obtinguts han estat considerats com a vàlids per part dels experts. Les publicacions relatives a la caracterització del domini són tres: dues que descriuen la caracterització de les dades, (Nicolas et al., 2008a) i (Nicolas et al., 2008b), i una tercera que fa referència a l'extracció de patrons (Vernet et al., 2008). Aquestes publicacions s'expliciten a continuació:

- R. Nicolas, E. Golobardes, A. Fornells, S. Puig, C. Carrera, and J. Malvehy, Identification of relevant knowledge for characterizing the melanoma domain. In *Advances in Soft Computing,* volume 49, pages 55-59. Springer Berlin-Heidelberg, 2008.
- R. Nicolas, E. Golobardes, A. Fornells, S. Puig, and J. Malvehy, Estudi de les característiques del domini del melanoma (code: 071208), *Technical Report.* EALS-URL, 2008.
- D. Vernet, R. Nicolas, E. Golobardes, A. Fornells, C. Garriga, S. Puig, and J. Malvehy, Pattern discovery in melanoma domain using partitional clustering. In *Frontiers in Artificial Intelligence and Applications*, volume 184, pages 323-330. IOS Press, 2008.

Arribats a aquest punt cal utilitzar les dades obtingudes del domini per a les diferents fases del sistema DERMA d'ajuda al diagnòstic. Aquesta aplicació és la que es descriurà en els capítols següents.

"Qui lluita pot perdre, qui no lluita ja ha perdut"
Bertold Brecht

3 Fita 2: Diagnòstics parcials no col·laboratius

Definit el domini per al càncer de melanoma, ja es poden començar els processos que utilitzen els conjunts de dades resultants. Les diferents etapes del procés permetran arribar a un sistema d'ajuda al diagnòstic global. El primer repte a afrontar és la classificació perquè el procés de diagnòstic mèdic es fonamenta en classificar nous pacients, per tant, en aquesta etapa els esforços es centren en la classificació. El procés analitzarà quines classes es consideren, quines dades s'utilitzen i com fer que el procés de classificació sigui més eficient a través de les millores proposades en la memòria de casos. Un cop assolits els objectius, el sistema de classificació resultant serà el fonament de les etapes posteriors. En pròxims capítols veurem com la combinació dels classificadors s'integra com a base del sistema d'ajuda global al diagnòstic.

3.1 Motivació i objectius

En un sistema d'ajuda al diagnòstic el primer pas és l'anàlisi de com classificar un nou cas a partir d'unes dades concretes. En el cas del càncer de melanoma aquest procés correspon a classificar un nou pacient segons les dades disponibles d'un expert o d'una prova concreta. El procés, seguint el protocol mèdic, es produeix establint similituds i diferències entre el nou cas i els pacients prèviament diagnosticats. Aquest model funcional s'ajusta al dels sistemes de CBR, que resolen nous problemes a través d'un procediment d'analogies basat en experiències representades per un conjunt de casos emmagatzemats en una memòria de casos. Per tant, CBR és capaç de justificar

les solucions obtingudes utilitzant analogies amb problemes anteriors, que és crucial per als experts.

L'adaptació dels sistemes classificadors CBR a la metodologia de diagnòstic de càncer de melanoma és el principal objectiu del capítol. Com hem descrit anteriorment, a DERMA cada conjunt de dades o mètode diagnòstic estarà representat pel seu propi classificador de manera que obtindrem tants sistemes especialitzats com tests o experts hi hagi involucrats en el diagnòstic. Aquests mòduls de classificació, basats en CBR, tindran un conjunt de característiques comunes: (1) la memòria de casos serà específica per a cada expert (o prova) i les dades que la conformen tindran les característiques pròpies d'aquell tipus de diagnòstic; (2) les fases de recuperació i adaptació seguiran el model habitual de CBR en què es recuperaran els pacients més similars al nou cas i la solució s'adaptarà a partir d'un procés de votació combinant els casos recuperats; (3) el procés de revisió el farà un expert del domini i, finalment, (4) la fase d'emmagatzematge guardarà, en una etapa inicial, els casos que s'hagin classificat que siguin significativament diferents dels prèviament emmagatzemats. En fases posteriors veurem com la classificació col·laborativa requereix guardar més informació durant el procés d'emmagatzematge que serà utilitzada per a l'adaptació.

Amb el sistema CBR implementat i testejat es va veure que la solució s'adaptava correctament al criteri demanat pels experts. Tot i així, i com veurem en detall al capítol 6, les ràtios de falsos negatius havien de millorar per satisfer totes les necessitats mèdiques i, per això, es va proposar una millora del procés de classificació a través d'un canvi en la memòria de casos utilitzant Aprenentatge de Mètriques de Distància / *Distance Metric Learning* (DML). Mitjançant aquest procés s'obté una memòria de casos organitzada d'una manera que en permet una millor recuperació (Nicolas et al., 2009a). La millora, com es veurà a la següent secció, es fonamenta en obtenir un nou conjunt de dades en què els pacients de la mateixa classe es mantinguin units i, al mateix temps, els casos de diferents classes es separin el màxim possible.

3.2 Millora de la memòria de casos

La millora en l'aplicació CBR es fonamenta en les mesures implementades anteriorment però sumant-hi la utilització de DML per obtenir millors resultats. Amb aquesta aportació volem aprendre una mètrica que mantingui a prop els punts de la mateixa classe i, al mateix temps, separar el màxim possible els punts de classes diferents. Amb el mètode aprenem una mètrica de distància global que minimitza la distància entre parells de punts del conjunt d'equivalència i que separa les dades que acompleixen les restriccions d'inequivalència. A partir d'aquests conjunts s'aconsegueix la projecció del conjunt que permet minimitzar el conjunt de distàncies.

L'aportació de DERMA es va testejar a través d'una plataforma prèvia anomenada *COllaborative MElanoma DIagnosis using CBR* (COMEDI-CBR) (Nicolas et al., 2009a). Com veurem en el capítol 6, l'aplicació de DML als diferents models experts permet millorar l'encert del sistema en termes de precisió i de falsos positius i negatius. Aquests valors són de vital importància en el diagnòstic mèdic ja que no es pot

permetre que un pacient amb la malaltia no sigui tractat correctament per un error en la seva classificació.

3.3 Conclusions i passos següents

En aquest capítol hem comprovat la necessitat de classificar els pacients de melanoma segons les seves característiques. El procés descrit ens permet definir un nevus com a maligne o benigne i com a melanocític o no melanocític. A més, s'ha establert un procediment de millora de la classificació a través de la modificació de la memòria de casos. El canvi es duu a terme utilitzant una projecció de les dades amb DML per optimitzar la classificació posterior. L'article publicat directament relacionat amb la millora dels sistemes parcials de CBR en càncer de melanoma (Nicolas et al., 2009a) és el següent:

- R. Nicolas, D. Vernet, E. Golobardes, A. Fornells, F. de la Torre, and S. Puig, Distance metric learning in a collaborative melanoma diagnosis system with Case-Based Reasoning. In *Proceedings of the 14th United Kingdom Workshop on Case-Based Reasoning at The 29th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence, CMS Press, University of Greenwich, UK,* pages 58-66. 2009.

Un cop establerta la classificació bàsica de les dades, el següent pas és aconseguir que sigui col·laborativa utilitzant les classificacions de diferents fonts de dades. L'objectiu es descriu en el capítol següent.

"Copia ciborum, subtilitas impeditur"

Lucius Annæus Seneca

4

Fita 3: Diagnòstic global col·laboratiu

El capítol anterior ens ha permès descriure els diagnòstics parcials en càncer de melanoma. Aquests sistemes independents tracten el diagnòstic d'un expert, d'una prova o d'un tipus de classificació. Seguint les indicacions del protocol de diagnòstic utilitzat pels experts mèdics constatem que, a partir de les classificacions, hem d'establir un patró global i col·laboratiu. Si ho fem d'aquesta manera podem aconseguir un diagnòstic únic que aglutini els criteris, les proves i les classificacions dels diferents especialistes. En aquest capítol veurem la descripció detallada d'allò que ha motivat el diagnòstic col·laboratiu, també s'analitzarà com s'ha dut a terme aquesta col·laboració a través del protocol mèdic. I s'introduiran les millores aportades a la col·laboració bàsica. Les millores es centren en afinar el procés de col·laboració a través de l'ús de coneixement del domini. La informació es recollirà en regles extretes de les dades per discernir entre l'opinió dels diversos subsistemes especialitzats. L'aportació de diagnòstic col·laboratiu permetrà assolir un dels objectius principals sol·licitats pels experts del domini: aglutinar en una de sola les decisions parcials.

4.1 Motivació i objectius

Com hem vist al capítol 2, sobre la caracterització del domini, en el diagnòstic del càncer de melanoma s'hi impliquen diferents especialistes i mètodes i, per tant, a l'hora de diagnosticar-lo s'utilitzen diferents fonts de dades. Això fa que un enfocament col·laboratiu sigui adequat per al diagnòstic del càncer de melanoma.

S'han publicat diversos treballs que utilitzen sistemes de col·laboració per permetre una millora en algorismes coneguts. Tot i que els sistemes de col·laboració generals

s'han utilitzat a bastament (Ontañón, 2008) en la classificació del melanoma s'han de considerar les característiques específiques que comporten la utilització dels coneixements mèdics. El protocol de col·laboració ha de seguir l'utilitzat pels experts en melanoma, com si es tractés d'una formalització algorísmica del seu procediment. El mètode és la combinació de diferents decisions de sistemes parcials per construir una solució més fiable, com ja s'ha fet en altres problemes (Melville and Mooney, 2004). En el nostre cas estem treballant amb dos punts de vista diferents (confocal i dermatoscòpic) dels quals seleccionem la millor classificació en funció de diferents criteris (Nicolas et al., 2008c) i, per tant, les característiques concretes del domini (Nicolas et al., 2008a) fan que sigui necessari emprar un mètode diferent del general.

Seguint les característiques del protocol mèdic i la seva aplicació dins del sistema DERMA podem resumir les seves funcionalitats amb el procés que detallarem a continuació. A partir d'un nou cas, format per dades d'imatges confocals i dermatoscòpiques, s'inicia el camí fent dues classificacions parcials, cadascuna de les quals es fa a través d'un mòdul de CBR especialitzat en aquest tipus de dades. El resultat dels diagnòstics parcials indica la predicció de melanociticitat i malignitat del nou cas. La segona capa del sistema estableix la col·laboració entre mòduls, per aconseguir generar una solució final unificada segons el protocol mèdic. El procediment concret de col·laboració el descriurem a la secció 4.2.

Amb tot, tenint en compte que els objectius principals són la millora de la classificació i la reducció màxima de falsos negatius, el sistema introdueix un mòdul de regles per garantir-ne la fiabilitat en situacions excepcionals causades per singularitats de les dades. Es tracta d'una segona capa del sistema de col·laboració (Nicolas et al., 2009b) en què un mòdul de regles preprocessa les dades d'entrada i genera un conjunt d'informació per ajudar el classificador.

A continuació, un cop descrit el marc global col·laboratiu, s'expliquen en detall les dues fases de la funcionalitat de DERMA: el sistema col·laboratiu bàsic seguint el protocol mèdic i l'ús de regles per millorar-ne la combinació.

4.2 Col·laboració seguint el protocol mèdic

L'objectiu fonamental d'aquest model expert (Nicolas et al., 2008c) és el desenvolupament d'un marc col·laboratiu per ajudar els experts en el diagnòstic mèdic utilitzant les dades diagnòstiques més destacades. Per fer-ho possible es volen combinar les decisions dels diferents experts i, per tant, dels diferents sistemes CBR independents. Per assolir el propòsit es va presentar una plataforma inicial de DERMA, publicada amb el nom de *MElanoma DIagnosis Based on Ensembles* (MEDIBE), que segueix el protocol mèdic de decisió per proposar un diagnòstic. El patró seguit pels metges (vegeu la figura 4.1) es fonamenta en dues etapes: la primera analitza si una nova mostra (imatge d'un nevus) és o no melanocítica i en la segona s'estudia la seva malignitat per extreure el diagnòstic final a partir de la combinació de les dues decisions. Tanmateix, cadascuna de les dues etapes té dues decisions parcials: la de l'expert en la imatge confocal i la del que ho és en dermatoscòpia. Les dues opinions són les que cal unificar

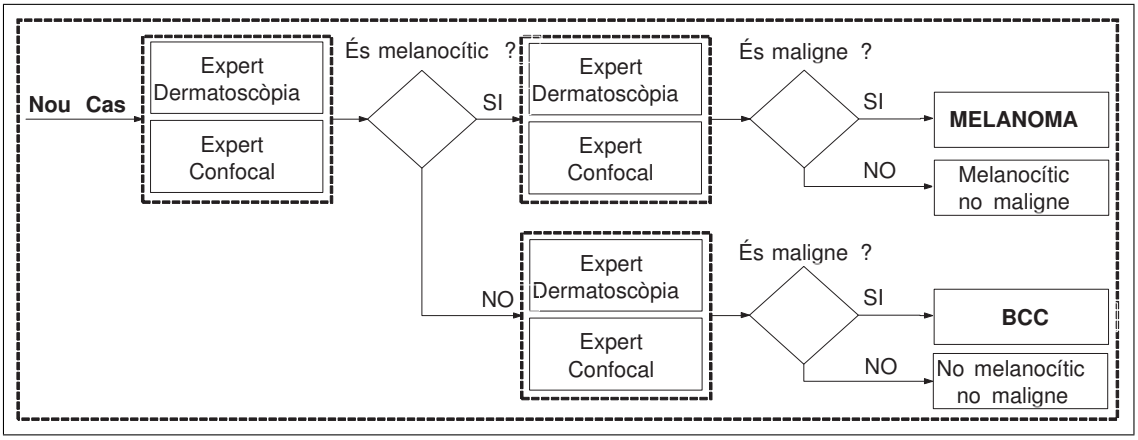

Figura 4.1: Protocol de diagnòstic mèdic seguit per experts en melanoma amb dades de dermatoscòpia i confocals. S'hi descriu el procés des de l'entrada d'un nou cas fins a la proposta de diagnòstic. L'esquema mostra els dos passos del diagnòstic amb la classificació amb criteri melanocític seguida de la classificació de malignitat.

per emetre'n una de conjunta.

El procés seguit per la plataforma d'ajuda al diagnòstic s'ajusta a les característiques del protocol mèdic, substituint els punts de combinació de decisions dels experts per un mòdul col·laboratiu. Disposem, doncs, de dos blocs CBR independents, en què cadascun té una memòria de casos específica. El nou cas és analitzat pels dos sistemes que emeten el seu pronòstic i les decisions parcials s'unifiquen a través del mòdul de combinació que ens dirà si el nou cas és positiu o negatiu. Aquest mòdul compara els resultats parcials recuperats amb el cas d'estudi per determinar quant de diferents (distants) són aquests casos. A partir de la distància obtinguda s'estableix si s'accepta el diagnòstic confocal o el dermatoscòpic a través d'un valor llindar.

En resum, el més destacable és que el sistema emula el protocol mèdic utilitzant com a experts dos sistemes CBR dels quals en combina les sortides. Els dos sistemes experts utilitzen una CM independent amb casos prèviament diagnosticats a partir d'estudis dermatoscòpics i confocals, respectivament. Els dos mòduls CBR són totalment independents i al final del procés emeten un diagnòstic propi, que serà unificat pel sistema de combinació. La decisió final, unificada, es genera seguint l'opinió dels experts en medicina que fixen el valor llindar d'acceptació donant més pes al criteri confocal (Nicolas et al., 2008c). A continuació veurem com es pot millorar el procés a través de regles de preprocessament.

4.3 Millora de la combinació fent servir regles

A la segona aportació inclosa al sistema DERMA, i implementada inicialment en la plataforma *Rule Based MElanoma DIagnosis Based on Ensembles* (RB-MEDIBE)

(Nicolas et al., 2009b), s'ha afegit un mòdul basat en l'anàlisi de les dades d'entrenament per obtenir un conjunt de regles que resumeixen la complexitat de les dades de la memòria de casos. L'objectiu principal és representar les llacunes existents a l'espai de dades sense informació associada per assessorar el classificador. Quan una regla garanteix amb una alta fiabilitat una classificació correcta, el sistema envia una alerta al classificador si detecta les característiques en les dades d'entrada. D'aquesta manera, els dos mòduls CBR segueixen el mateix procediment que hem vist fins ara però el mòdul de combinació queda condicionat a un bloc de regles.

Atenent al criteri mèdic, en el domini del melanoma és habitual que valors concrets de determinats atributs fixin la classificació final. En aquest cas la classe és independentment del valor que prenguin els altres atributs. Utilitzant aquest coneixement s'han analitzat els espais de dades que tenen el comportament anteriorment citat i per a cadascun d'ells s'ha generat una o més regles (atenent a si el conjunt fixa un o més valors de classe). L'algorisme de generació de regles fa servir tots els possibles valors de cada atribut d'entrada. Sempre que el sistema troba un valor d'un atribut al que li correspon una classe única per a tots els casos d'entrenament el mòdul de regles genera una nova regla. Un dels avantatges en l'àmbit mèdic és que els atributs estan ben delimitats, per tant, és difícil trobar un nou cas amb un valor diferent dels predefinits en el domini.

El mòdul de regles està connectat al sistema inicial a través de la combinació: utilitza el millor dels casos recuperats, com en el cas sense preprocés de les dades, però afegint-hi la informació proporcionada per les regles per poder-lo ponderar. Això permet la possibilitat de seleccionar de forma automàtica el millor cas i aconseguir un diagnòstic més acurat.

El procés de generació del conjunt de regles i la seva posterior utilització com a complement al mòdul de col·laboració permet una millora del sistema gràcies a la possibilitat d'establir un diagnòstic més acurat i, al mateix temps, d'ajustar-se més al protocol utilitzat pels experts que també valoren l'aparició repetida d'atributs concrets per determinar el diagnòstic d'un nou cas.

4.4 Conclusions i passos següents

En aquest capítol hem observat la necessitat que DERMA passi a ser un sistema col·laboratiu. La decisió s'ha fonamentat a partir de la diversitat de dades, experts i mètodes que s'utilitzen en el diagnòstic mèdic del càncer de melanoma. A més, hem detallat les diverses opcions explorades per a la creació del sistema col·laboratiu i la seva implementació final. El primer model proposat fa una combinació seguint de manera estricta el protocol mèdic de col·laboració de tècniques. També hem presentat una millora a la col·laboració basada en regles per millorar la precisió dels resultats. El procés s'ha basat en el criteri dels experts d'establir regles a partir de la presència de determinats atributs d'un cas. A l'apartat 6.3 es detalla quin és l'impacte d'aquesta fase en la classificació dels nevus utilitzant la millora de la combinació amb regles i sense ella. Les publicacions d'aquest àmbit són dues, la primera de l'aplicació MEDIBE que

tracta els sistemes col·laboratius per a càncer de melanoma amb l'aportació general (Nicolas et al., 2008c), i la segona anomenada RB-MEDIBE que descriu l'optimització del sistema utilitzant regles (Nicolas et al., 2009b). A continuació es citen els articles relacionats amb les aportacions:

1. R. Nicolas, E. Golobardes, A. Fornells, S. Segura, S. Puig, C. Carrera, J. Palou, and J. Malvehy, Using Ensemble-Based Reasoning to help experts in melanoma diagnosis. In *Frontiers in Artificial Intelligence and Applications*, volume 184, pages 178-185. IOS Press, 2008.

2. R. Nicolas, D. Vernet, E. Golobardes, A. Fornells, S. Puig, and J. Malvehy, Improving the combination of CBR systems with preprocessing rules in melanoma domain. In *Workshop Proceedings of the 8th International Conference on Case-Based Reasoning, Seattle, WA, USA,* pages 225-234. 2009.

Assolit l'objectiu principal del capítol, que és el diagnòstic col·laboratiu, el següent pas en la plataforma DERMA és la utilització de conjunts de dades més rics, com són els multietiqueta, un canvi que permetrà fer una classificació de més d'una característica alhora. La necessitat que hem vist en la tipologia de les dades i el diagnòstic es detallarà en el capítol següent.

"Qui perd els seus orígens, perd la seva identitat"
Joan Salvat-Papasseit

5 Fita 4: Diagnòstic multietiqueta

En el capítol anterior hem vist com la capacitat col·laborativa és d'ajuda en el diagnòstic de càncer de melanoma, però també hem comprovat que en determinats casos la col·laboració no preveu tota la riquesa de les dades. Fins a aquest punt la classificació realitzada ha considerat una única etiqueta cada vegada, per això, un cop assolida la fita de l'ajuda al diagnòstic col·laboratiu, l'evolució natural del sistema és l'ús de dades amb més d'una etiqueta. El canvi permetrà aprofitar totes les característiques del coneixement del domini. En aquest capítol es descriu què ha motivat el pas d'un sistema d'una sola etiqueta a un de múltiples. El canvi es fonamenta en el problema que suposa la necessitat de discernir entre diferents tipus de nevus a partir d'un únic cas d'entrada. A més, es detallen les característiques de la millora. La principal diferència en relació a la versió amb una sola etiqueta de DERMA està en la reutilització dels casos. En aquest sentit es proposen dues maneres de donar solució al problema: l'una a través d'un procés probabilístic i l'altra, afegint-hi l'experiència. Amb aquesta aportació al diagnòstic de melanoma amb dades multietiqueta s'aconsegueix que el sistema tracti amb els diferents tipus de dades proposats pels experts del domini.

5.1 Motivació i objectius

De la integració del protocol mèdic a DERMA en destaca un aspecte interessant: el diagnòstic final s'obté a partir de la consideració de diferents patrons de classificació. Aquesta característica emprada pels experts mèdics és el principi fonamental dels

problemes de classificació Multietiqueta (ME), en què la classificació final està dividida en parts per evitar la pèrdua d'informació. Segons les característiques del domini, els experts mèdics i els sistemes multietiqueta es conclou que el càncer de melanoma està millor representat fent servir dues classes no disjuntives, que són melanocítiques i malignes, en lloc d'utilitzar una sola classe. I aquesta conclusió és la que motiva el pas del sistema DERMA a multietiqueta.

A continuació veurem en detall quin ha estat el procés d'adaptació a sistemes multietiqueta del sistema DERMA d'una sola etiqueta que s'havia descrit fins ara.

5.2 Raonament analògic multietiqueta

L'adaptació de DERMA a les característiques ME s'ha fet segons les condicions de la família Mètodes d'Adaptació d'Algorismes / *Algorithm Adaptation Methods* (AAM). La nostra proposta planteja un sistema de CBR amb característiques multietiqueta fonamentades en l'adaptació del sistema KNN que fa *Multi-Label K-nearest neighbor* (MlKnn). Si bé Zhang and Zhou (2005) adapten KNN al seu funcionament ME, DERMA ho fa amb CBR.

Les quatre etapes del cicle de CBR s'han adaptat a les característiques multietiqueta de la manera següent: (1) la recuperació segueix el mateix patró que l'algorisme regular de CBR. El sistema selecciona els casos més similars d'acord amb el valor obtingut a partir d'una funció de similitud, és a dir, es recuperen d'entre la memòria de casos aquells que són més similars al nou; (2) la reutilització s'adapta al funcionament ME d'una manera probabilística. El mètode es basa en comptar les ocurrències de cada etiqueta i considerar que és positiva si més de la meitat dels casos recuperats tenen l'etiqueta amb valor positiu. El procés és similar al proposat per a l'algorisme MlKnn. La fase de reutilització es millora considerant l'experiència de cada cas recuperat (obtinguda a través de la fase de manteniment) per ponderar apropiadament els casos recuperats; (3) la fase de manteniment ajusta la informació assignant als casos recuperats un indicador de l'èxit o el fracàs de la classificació i, finalment, (4) la revisió, com en una sola etiqueta, és realitzada per un metge especialista.

En primer lloc la proposta centra l'esforç en la recuperació i reutilització de CBR. L'adaptació de les fases permetrà mantenir o millorar la precisió aconseguida per altres mètodes ME tot adaptant-los al nostre problema concret. De la mateixa manera, s'adapten també les altres fases de CBR per tancar el cicle complet. Per a la fase de reutilització es proposen dos enfocaments diferents: la primera opció realitza la classificació final a través d'un procés de votació en què tots els casos recuperats tenen el mateix pes, i en la segona opció, s'hi afegeix el concepte d'experiència per ajustar el pes dels diferents casos recuperats. En els apartats següents es detallen els algorismes de classificació proposats amb els diferents tipus de recuperació.

Cal destacar que l'aportació s'ha fet des de dos punts de vista, un no basat en el diagnòstic mèdic sinó aplicable a dades generals que s'ha anomenat *Multi-Label Case-Based Reasoning* (MlCBR) (Nicolas et al., 2013b), i un altre aplicat al càncer de

melanoma integrat dins del sistema DERMA (Nicolas et al., 2013a).

5.2.1 Reutilització probabilística

Com hem descrit, tant DERMA multietiqueta com MlCBR utilitzen com a base un sistema CBR que adapta les seves fases per tractar amb més d'una etiqueta. En la primera fase s'utilitza l'algorisme de Reutilització Probabilística / *Probabilistic Reuse* (PR) (Nicolas et al., 2013b), que presenta variacions probabilístiques en l'etapa de reutilització clàssica del raonament basat en casos per adaptar-se a la classificació ME. La plataforma recupera els k millors casos i els explora amb cada etiqueta per actualitzar el seu percentatge d'aparicions. Finalment, el sistema decideix si l'etiqueta és positiva o no a través d'un valor llindar. Com en el cas amb una sola etiqueta el valor llindar es fixa seguint el criteri mèdic.

5.2.2 Reutilització probabilística amb experiència

El segon pas proposa una millora de la fase de reutilització amb l'ús de l'experiència. Aquesta fase, anomenada Reutilització Probabilística amb Experiència / *Probabilistic Reuse based on Experience* (PRE), afegeix un valor positiu a la classificació recuperada si és correcta i un de negatiu si no ho és. La informació s'utilitzarà en altres classificacions per tenir més o menys en consideració un determinat cas recuperat. El procés té dues parts: en primer lloc, la memòria de casos té un valor associat a cada cas que té en compte l'experiència obtinguda pel cas en les classificacions anteriors, un valor que s'actualitza després de cada iteració de l'algorisme de classificació d'un nou pacient, i en segon lloc, durant el procés de classificació d'un nou cas, l'algorisme multiplica el percentatge d'aparició de cada etiqueta d'un cas per l'experiència associada al cas al qual pertany. Aquest procés de ponderació es realitza amb la finalitat de tenir en major o menor consideració els casos recuperats durant el procés de reutilització.

5.3 Conclusions i passos següents

En aquest capítol hem vist la utilitat dels sistemes ME i com poden ajudar al diagnòstic de càncer de melanoma. La classificació multietiqueta ens permet una millor descripció del problema i perdre menys informació durant el procés de classificació, dues característiques extremadament importants en un problema mèdic. Per afrontar la problemàtica s'ha proposat una adaptació del sistema CBR per utilitzar-lo amb dades de més d'una etiqueta i amb aquesta quarta fita queda completada la plataforma DERMA. Una primera fase aborda el problema des de l'adaptació de la fase de reutilització per poder tractar amb més d'una etiqueta a l'hora de donar un pronòstic. En una segona fase s'afegeix experiència al procés fent que durant l'emmagatzematge es guardi informació relativa a l'èxit de la classificació. La bondat d'un cas a partir dels èxits i fracassos previs condicionarà la tria d'una o altra etiqueta. En el present

capítol les publicacions es divideixen en dos grups. En primer lloc la publicació Nicolas et al. (2013b) està directament relacionada amb els sistemes multietiqueta descrits en aquest capítol i, en concret, amb el sistema MlCBR que centra els seus esforços en la classificació multietiqueta general. La referència es cita a continuació:

1. R. Nicolas, A. Sancho-Asensio, E. Golobardes, A. Fornells, and A. Orriols-Puig, Multi-label classification based on analog reasoning, *Expert Systems With Applications Journal,* 40:5924-5931. 2013.

En segon lloc cal destacar també dues publicacions més, no relacionades directament amb la classificació ME però que sí la inclouen. Fan referència a l'aplicació DERMA completa (Nicolas et al., 2013a, 2014) on es descriu l'aplicació al domini del càncer de melanoma de totes les fites descrites. Es troben referenciades a continuació:

1. R. Nicolas, A. Fornells, E. Golobardes, G. Corral, S. Puig, and J. Malvehy. Melanoma diagnosis based on collaborative multi-label reasoning. In *Frontiers in Artificial Intelligence and Applications*, pages 283-292. IOS Press, 2013. DOI: 10.3233/978-1-61499-320-9-283.
2. R. Nicolas, A. Fornells, E. Golobardes, G. Corral, S. Puig, and J. Malvehy. DERMA: A melanoma diagnosis platform based on collaborative multi-label analog reasoning. *The Scientific World Journal*, 2014.

Per últim, com es veurà a l'apartat 6.6 una de les principals línies de futur d'aquesta tesi és la utilització en altres dominis d'aplicació de les tècniques aplicades al càncer de melanoma. Un d'aquests dominis és el de l'educació en el que ja s'hi està treballant en conjunt amb altres membres del grup de recerca. És per això que una tercera publicació es centra en l'aplicació de les tècniques presentades en aquesta tesi en l'àmbit de l'educació (Vernet et al., 2010) que es cita a continuació.

1. D. Vernet, R. Nicolas, E. Golobardes, A. Fornells, and A. Garcia-Piquer, Intelligent tutoring system framework for the acquisition of knowledge and competences. In *40th ASEE/IEEE Frontiers in Education Conference*, pages 111-112. IEEE, 2010.

Amb el sistema descrit completament a través dels darrers capítols ja podem iniciar el procés d'experimentació que es descriurà en el capítol següent.

"Qui no sàpiga conversar, sigui un individu, una col·lectivitat, un estat, no mereix sinó el menyspreu"

Salvador Espriu

6

Anàlisi de resultats, conclusions i línies de futur

En els capítols previs hem vist la descripció de les diferents aportacions del sistema DERMA. Les fites analitzades s'han descrit en termes de context, motivació i característiques de la solució proposada, per tant, manca l'estudi dels resultats obtinguts per a cadascuna de les fases d'aquest sistema. El primer objectiu del capítol és analitzar la metodologia d'experimentació utilitzada, els conjunts de dades i els resultats obtinguts. El procés experimental es dividirà en tres etapes: una primera destinada a la fase prèvia a DERMA, d'extracció de característiques del domini, una segona a l'estudi dels resultats de DERMA amb les dades mèdiques i una darrera amb dades genèriques per testejar el funcionament del sistema amb altres dominis. Els diferents blocs experimentals, a més, vénen acompanyats d'una discussió dels resultats obtinguts. Un cop analitzats els resultats, cal fer una anàlisi més qualitativa de la tasca realitzada. A l'apartat de conclusions veurem què aporta DERMA, i el conjunt de la tesi, des d'una òptica menys centrada en els resultats. L'estudi permetrà verificar que aquesta eina és útil com a complement a la decisió. Per últim, descriurem quines són les línies de futur a tractar a partir d'ara, que es centren en l'ampliació de les dades i experts que interactuen amb DERMA i l'extensió de les tècniques aplicades a altres dominis.

6.1 Motivació i objectius

Com hem vist a través de les aportacions d'aquesta tesi, l'objectiu final és proposar una eina capaç d'ajudar els experts en càncer de melanoma a fer el seu diagnòstic. La fita global s'ha aconseguit mitjançant la plataforma DERMA, que proposa un marc per a les diferents tècniques i mètodes d'ajuda que hem treballat. Tot i així, l'experimentació realitzada en aquesta tesi no es centra només en el sistema sinó que s'aplica també a fases anteriors i posteriors a la plataforma i, per això, no s'ha fet en un únic bloc sinó que s'ha dividit en tres parts. En el primer cas s'analitzen els resultats previs a DERMA i, per tant, els que són propis de la caracterització i preparació del domini de dades. En segon lloc, es tracten els resultats propis de DERMA amb les dades mèdiques aportades pels experts. Per últim, el tercer bloc d'experimentació, mostra els resultats d'utilitzar les tècniques aglutinades per DERMA aplicades a altres dominis. El tercer grup ens permet donar una visió global de la utilitat del sistema. A continuació veurem els detalls de cadascuna d'aquestes parts.

6.2 Resultats de la caracterització del domini

El primer bloc d'experimentació és el referent als resultats quantitatius de la primera fita: caracterització del domini. Per tant, en aquesta secció veurem l'experimentació realitzada en l'extracció de patrons en el càncer de melanoma. Aquest procés és el que es fa a través de la clusterització per aconseguir patrons i establir la tipologia d'un cas. La caracterització del domini pròpiament dita, que és de caràcter qualitatiu, està definida en el capítol 2. La fase experimental realitza una anàlisi de 2 a 20 classes (clústers) amb 10 llavors per cadascuna. Els valors s'han fixat d'acord amb els experts del domini que consideren que 20 classes són suficients considerant el nombre d'instàncies existent (Vernet et al., 2008). Amb els resultats, s'ha creat un nou conjunt de dades per a cada llavor i nombre de classes, que utilitza les mateixes instàncies que el conjunt inicial però etiquetades de nou amb l'atribut de classe basat en el clúster en què ha estat assignada.

Un cop establerts els nous atributs de classe, s'han validat com a atributs de classificació. El procés s'ha realitzat a través de l'estudi del percentatge d'error de classificació amb CBR. D'aquesta manera s'ha utilitzat un sistema de CBR amb els casos etiquetats utilitzant les noves classes com a memòria de casos i el resultat obtingut ha estat el percentatge d'encerts de classificació utilitzant les noves dades. La figura 6.1 resumeix els resultats del procés: a l'eix x s'hi mostra el nombre de clústers de la classificació (valor de k a l'algorisme *k-means*) i a l'y el percentatge d'encerts per aquell valor de k. La gràfica mostra el valor mitjà d'encerts de les diferents llavors de cada classificació.

Analitzant els resultats observem: (1) que els patrons de classificació generats aconsegueixen una precisió mínima d'un 50 per cent i màxima del 85 per cent, (2) que els límits superior i inferior de la precisió corresponen, respectivament, al major i menor nombre de clústers, (3) que la tendència és decreixent excepte en el cas de 9 clústers on hi ha un pic positiu d'un 10 per cent en el percentatge d'encerts respecte a les

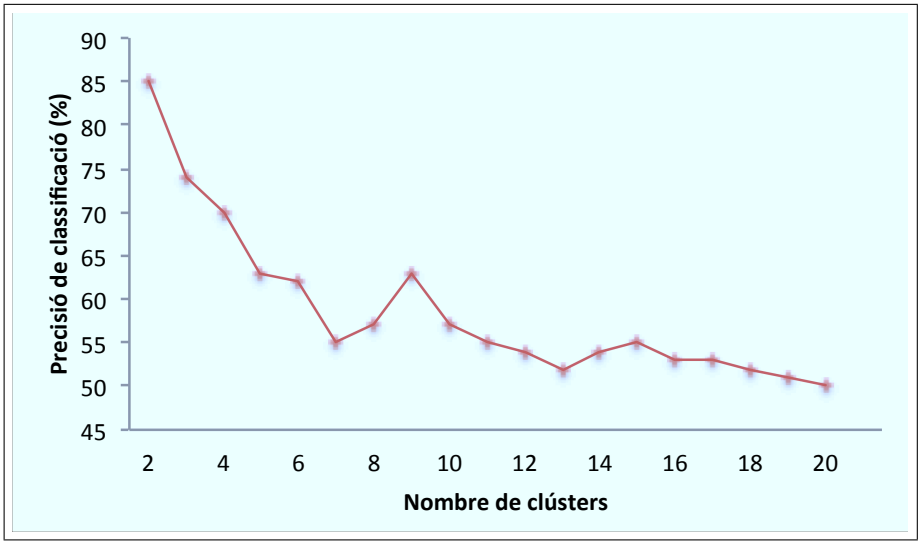

Figura 6.1: Resultats de precisió en la classificació a partir dels clústers artificials. L'eix x fa referència al valor de k de l'algorisme k-means i l'eix y representa la precisió de classificació. La gràfica mostra l'evolució de la classificació amb els diferents nombres de clústers.

altres combinacions de k properes. En resum, l'anàlisi quantitativa remarca quatre idees fonamentals: (1) els resultats de classificació són millors amb pocs clústers, (2) hi ha un màxim local amb 9 clústers, (3) es denota una tendència decreixent i (4) una recomanació de criteri de classificació és utilitzar 9 clústers. El darrer punt lliga amb la classificació mèdica actual que utilitza 8 classes. A escala qualitativa s'ha fet una definició de les característiques dels clústers obtinguts en la configuració $k = 9$ per tal que els experts del domini l'avaluïn en comparació amb la classificació amb les 8 classes que feien servir fins ara. Els especialistes de l'HCPB apunten que la nova configuració aporta interessants grups de dades i que s'ajusta a les necessitats mèdiques.

6.3 Resultats amb dades de càncer de melanoma

La col·laboració amb experts de l'HCPB ha jugat un paper important en totes les decisions adoptades per dissenyar DERMA i també en la preparació i estudi dels experiments realitzats amb la plataforma. Considerant la col·laboració i les dades de les quals es disposa a la base de dades de melanoma, l'experimentació s'ha dividit en dos grups: el primer bloc cobreix l'ús de dades dermatoscòpiques i confocals per provar l'enfocament de col·laboració dels subsistemes CBR amb els dos tipus de dades, i el segon treballa també amb les dades d'imatges de nevus però amb múltiples etiquetes per provar DERMA al complet. A la següent secció veurem el detall de les dades utilitzades per a l'experimentació. Així mateix, abans de descriure les fites aconseguides per la plataforma i els resultats obtinguts, es descriurà el marc experimental utilitzat.

6.3.1 Banc de proves

Com hem vist al capítol 2, la caracterització del domini ens ha permès crear un model que tingui en compte els diferents experts clínics, tipus d'informació i proves per definir l'estructura ideal del conjunt de dades. Dins d'aquest model, i atenent el criteri dels experts de l'HCPB, les tècniques més utilitzades per obtenir informació dels teixits són: les imatges dermoscòpiques i les imatges confocals. El conjunt de dades utilitzat per a aquest apartat de l'experimentació està format per 150 casos de lesions sospitoses. Per a tots els casos tenim la informació relacionada amb les imatges confocals i dermatoscòpiques així com la histologia que corrobora el diagnòstic. Segons les consideracions dels experts que han creat el conjunt de dades, aquest és representatiu del domini. En termes mèdics es tracta d'una memòria de casos idònia per a aquest estudi. Detallant els casos, la informació dermatoscòpica té 41 camps i la microscòpia confocal, per la seva major resolució, contribueix amb dades de 83 atributs diferents. Les característiques concretes dels atributs així com el tractament de cada tipus de dades queden recollits a (Nicolas et al., 2008c). Les dades s'han configurat de dues maneres: (1) com un conjunt de dades amb una sola etiqueta per a cada procés de classificació, que es basa en els atributs apropiats per classificar una classe, i (2) un conjunt de casos amb totes les etiquetes disponibles i que explota totes les propietats de les dades per classificar les diverses classes possibles alhora.

6.3.2 Resultats amb dades d'una etiqueta i la seva discussió

El procés experimental s'ha dut a terme seguint les finalitats mèdiques i això ens ha permès analitzar els resultats més interessants per als experts com són els falsos positius. L'experimentació amb les dades mèdiques d'imatges confocals i dermatoscòpiques ha testejat DERMA utilitzant diferents configuracions i analitzant la sensibilitat, especificitat i precisió del sistema. L'estudi considera dos subsistemes independents de CBR amb tres configuracions: (1) una de bàsica sense aplicar cap millora, (2) una amb l'ús de regles obtingudes a través de l'algorisme de preprocessament i (3) una altra amb l'aplicació de DML a les dades originals com s'ha exposat al capítol 3. A més, s'ha provat l'exactitud dels dos sistemes independents (un per a dades dermatoscòpiques i un altre per a confocals) per comparar-ne els resultats amb els del sistema col·laboratiu.

Els resultats obtinguts en classificació amb una sola etiqueta es resumeixen en les taules 6.1, 6.2, i 6.3 que, respectivament, mostren sensibilitat, especificitat i precisió de la classificació de noves lesions. L'experimentació es realitza per als diferents esquemes definits a DERMA així com per als sistemes CBR independents. Els resultats mostren les quatre classes possibles considerades pel protocol mèdic. La significació estadística dels resultats es representa amb un $\uparrow$ si és significativament millor i - si és equivalent.

La discussió dels resultats posa de manifest: (1) que les diferents capes de DERMA aconsegueixen millors resultats de classificació que els sistemes no col·laboratius independents o la combinació plana de DERMA, (2) que l'ús de la combinació dels dos tipus d'imatges amb l'ajuda de regles de preprocessament obté un augment important de les taxes de sensibilitat i especificitat (els resultats més importants per als experts

	No maligne		Maligne	
	Melanocític	**No melanocític**	**Melanocític** (Melanoma)	**No melanocític** (BCC)
CBR Dermatoscòpic	75%	80%	73%	81%
CBR Confocal	74% -	92% ↑	73% -	92% ↑
Col·laboratiu	80% ↑	94% -	70% -	92% -
Col·laboratiu+Regles	95% ↑	95% -	81% ↑	92% -
Col·laboratiu+Regles+DML	100% ↑	100% ↑	100% ↑	100% ↑

Taula 6.1: Resultats de sensibilitat obtinguts en classificació melanocítica, de melanoma i de BCC, amb una etiqueta, a través dels diferents reptes de DERMA.

	No maligne		Maligne	
	Melanocític	**No melanocític**	**Melanocític** (Melanoma)	**No melanocític** (BCC)
CBR Dermatoscòpic	95%	99%	92%	96%
CBR Confocal	99% ↑	98% -	96% ↑	95% -
Col·laboratiu	96% -	97% -	95% -	96% -
Col·laboratiu+Regles	99% -	99% -	98% ↑	100% ↑
Col·laboratiu+Regles+DML	100% -	100% -	100% ↑	100% -

Taula 6.2: Resultats d'especificitat obtinguts en classificació melanocítica, de melanoma i de BCC, amb una etiqueta, a través dels diferents reptes de DERMA.

	No maligne		Maligne	
	Melanocític	**No melanocític**	**Melanocític** (Melanoma)	**No melanocític** (BCC)
CBR Dermatoscòpic	90%	96%	87%	96%
CBR Confocal	88% -	95% -	90% ↑	95% -
Col·laboratiu	92% ↑	94% -	89% -	95% -
Col·laboratiu+Regles	98% ↑	99% ↑	94% ↑	99% ↑
Col·laboratiu+Regles+DML	100% ↑	100% ↑	100% ↑	100% ↑

Taula 6.3: Resultats de precisió obtinguts en classificació melanocítica, de melanoma i de BCC, amb una etiqueta, a través dels diferents reptes de DERMA.

mèdics), (3) que l'ús de la tècnica de DML, per classificar millor els nous casos, aconsegueix el desig dels experts mèdics, és a dir, evitar falsos negatius permetent un millor diagnòstic dels pacients, (4) que és més fàcil classificar els casos no malignes que els malignes i que aquest fet està relacionat amb la utilització d'un conjunt de dades del món real que s'ajusta a les característiques de la població en què hi ha més pacients amb nevus no malignes que amb malignes, (5) que els casos no melanocítics estan més ben classificats a causa de les característiques del problema i que, finalment, (6) els resultats del t-test mostren que cap millora de DERMA té un efecte significativament negatiu.

	Sensibilitat	Especificitat	Precisió
Multilabel Dermatoscòpic	86%	89%	92%
Multilabel Confocal	93% ↑	97% ↑	96% ↑
Multilabel Col·laboratiu	91% -	96% -	95% -
Multilabel Col·laboratiu+Regles	94% ↑	99% ↑	98% ↑
Multilabel Col·laboratiu+Regles +DML	100% ↑	100% -	100% ↑

TAULA 6.4: Resultats de sensibilitat, especificitat i precisió obtinguts, amb múltiples etiquetes, a través dels diferents reptes de DERMA.

6.3.3 Resultats amb dades multietiqueta i la seva discussió

L'experimentació amb dades multietiqueta ha seguit el mateix criteri que la d'una etiqueta sola. S'han utilitzat dades d'imatges confocals i dermatoscòpiques per analitzar la sensibilitat, especificitat i precisió de DERMA. Com en el cas anterior, es consideren dos subsistemes independents amb les mateixes tres configuracions. La diferència principal del procés és el conjunt de dades utilitzat, que ara és ME, i el mateix algorisme de DERMA que amb un sol procés de classificació dóna els resultats finals del pronòstic. Aquesta característica s'ha materialitzat en l'ús de dues classes en cada cas (melanocític i maligne) que ofereixen als experts tota la gamma de classificacions d'un nevus.

Els resultats obtinguts per DERMA es mostren a la taula 6.4 i mostren el percentatge d'èxits considerant les quatre possibles classificacions de nevus. La taula està formada pels resultats de la sensibilitat, especificitat i precisió obtinguts utilitzant un protocol no col·laboratiu (amb dades confocals i dades dermatoscòpiques respectivament), amb un patró de col·laboració amb la combinació de totes dues dades mitjançant combinació bàsica, amb regles de preprocés, i amb DML. Igual que en una sola etiqueta, la significança es representa amb un ↑ si el resultat és significativament millor i - si és equivalent.

Els resultats obtinguts posen de manifest que: (1) com en la classificació d'una sola etiqueta, les diferents capes de DERMA aconsegueixen millors resultats de classificació que els dels sistemes no col·laboratius independents, (2) els resultats utilitzant el sistema complet permeten l'èxit del diagnòstic en tots els casos. Tot i que aconseguim un cent per cent de precisió no estem davant d'un sistema infal·lible sinó davant d'un que sap com treballar amb les peculiaritats del domini, igual que els experts mèdics, (3) la classificació de totes les classes en un pas únic no perd cap mena d'informació, (4) cadascun dels increments de DERMA és positiu amb significació estadística o, almenys, equivalent en comparació amb el no-ús de la millora.

Addicionalment es va realitzar un estudi de DERMA utilitzant altres tipus de dades multietiqueta i comparant els resultats obtinguts amb els d'altres sistemes de classificació amb més d'una etiqueta (Nicolas et al., 2013b). L'estudi ens va permetre ajustar les característiques ME del sistema i corroborar que és un mètode competent en aquest tipus de classificació. A continuació es detallen els resultats obtinguts.

6.4 Resultats en altres dominis

Tenint en consideració la voluntat de corroborar la qualitat del mètode i de generalitzar l'ús del sistema multietiqueta amb CBR es va decidir fer un segon estudi amb dades de *benchmark* multietiqueta i amb dades sintètiques. D'aquesta manera es pretén veure l'adaptació del mètode a altres dominis. En aquest apartat es descriu l'experimentació amb altres conjunts de dades.

6.4.1 Banc de proves

Per a l'experimentació amb MlCBR, nom amb el qual s'ha publicat la versió no mèdica de DERMA, hem utilitzat deu conjunts de dades: tres són problemes reals utilitzats per la comunitat que treballa en algorismes de classificació ME i set són problemes sintètics generats amb un algorisme de generació de conjunts de dades proposat per nosaltres mateixos. L'elecció d'aquests conjunts de dades i la necessitat de proposar dades sintètiques està condicionada per les característiques del domini i la disponibilitat de dades ME. En contrast amb les tasques de classificació d'una sola etiqueta, en què hi ha una gran quantitat de conjunts de dades representatives per avaluar i comparar els diferents enfocaments algorísmics, aquest no és el cas del camp ME. La majoria dels treballs publicats (Avila et al., 2009) utilitzen tres conjunts de dades (*scene*, *emotion* i *yeast*) per comparar els diferents enfocaments. La generació del conjunt de dades sintètiques s'ha descrit a Nicolas et al. (2013b).

6.4.2 Resultats obtinguts i discussió

El present bloc d'experimentació té característiques diferents als anteriors. La metodologia experimental, els tests aplicats i els resultats mostrats són dissemblants. Els canvis els motiva que es tracta d'una plataforma genèrica i no aplicada al càncer de melanoma i, per això, hem d'aplicar l'experimentació no només per provar el sistema en el domini, sinó també per comparar-lo amb altres aplicacions del mateix tipus. De la mateixa manera s'hi apliquen els mateixos tests estadístics utilitzats per aquesta comunitat. Seguint aquest criteri, el primer pas ha estat la definició dels diferents paràmetres de la nostra proposta i a partir del procés experimental s'han fixat els valors adients de k i experiència (Nicolas et al., 2013b). En segon lloc, s'han analitzat els resultats obtinguts pels mètodes de referència en la classificació amb diverses etiquetes per comparar-los amb els de MlCBR. La comparació s'ha realitzat amb la finalitat de demostrar que MlCBR és competitiu en relació amb els algoritmes de referència en aquesta àrea.

Arribats a aquest punt, tenim tots els resultats de rendiment de les diferents plataformes de forma separada. A partir dels resultats obtinguts podem descriure la comparació global de MlCBR amb PR i PRE, i els algoritmes MlKnn i *RAndom k-labELsets* (RAkEL) en termes de precisió i significança estadística de la comparació. Els sistemes es classifiquen com es mostra a la taula 6.5, en què podem veure per

Algorisme	Rànquing	Posició
MlKnn	2,80	1
MlCBR (PRE k11 exp05)	2,90	2
MlCBR (PR k13)	2,95	3
MlCBR (PRE k11 exp03)	3,25	4
MlCBR (PR k09)	3,30	5
RAkEL	5,80	6

TAULA 6.5: Posició mitjana de Friedman i posició en el rànquing dels millors algoritmes.

	PR k09	PR k13	PRE k11 exp03	PRE k11 exp05	MlKnn	RAkEL
PR k09						
PR k13	+					
PRE k11 exp03	+	−				
PRE k11 exp05	+	+	+			
MlKnn	+	+	+	+		
RAkEL	⊖	⊖	⊖	⊖	⊖	

TAULA 6.6: Comparacions per parells dels mètodes d'aprenentatge per mitjà del procediment de Holm.

columnes el nom de l'algorisme provat, el seu valor d'ordenació i la posició obtinguda. Malgrat que el rànquing mostra com a millor el funcionament de MlKnn, davant de les millors configuracions de MlCBR amb PR i PRE, els resultats de significació estadística mostren que la diferència no és estadísticament significativa per $\alpha = 0,05$ ni per $\alpha = 0,10$. En canvi, no és el cas de RAkEL que, en referència als resultats estadístics, té diferències amb MlCBR (amb les dues etapes de reutilització) i MlKnn que són estadísticament significatives. Els resultats es relacionen amb la classificació, en la qual veiem que els resultats de precisió de RAkEL són significativament pitjors que els altres. A partir d'aquests resultats es pot concloure, en termes de precisió, que el nostre sistema (amb PR i PRE) és equivalent a altres plataformes competents en aquest àmbit, com MlKnn, i millor que RAkEL, un dels sistemes de referència per a la classificació ME.

Com a resum d'aquests resultats podem analitzar la taula 6.6, que fa una comparació per parelles dels mètodes d'aprenentatge per mitjà d'un procediment de Holm. Els resultats han de ser analitzats seguint la diagonal de la taula que mostra el símbol ⊖ si el mètode de la fila degrada significativament el mètode de la columna, a un nivell de significació de 0,05, el símbol + si es té un millor rendiment però amb cap diferència significativa, o el símbol − si, amb una diferència no significativa, es realitza pitjor. La comparació per parelles mostra gràficament que el rendiment de RAkEL és significativament pitjor en tots els casos, mentre que els resultats dels altres mètodes de prova són estadísticament equivalents. En afegit, a Nicolas et al. (2013b) es descriu la millora en termes computacionals del mètode multietiqueta MlCBR respecte als altres mètodes d'estudi.

6.5 Conclusions

Com hem vist, el càncer de melanoma és un problema creixent en la nostra societat. Les característiques de la malaltia i els diferents tipus de tècniques i professionals implicats en el diagnòstic fan que sigui important el disseny d'una plataforma que cobreixi la totalitat del problema. DERMA va néixer per aconseguir aquest objectiu a partir de les característiques extretes de l'anàlisi del domini. La plataforma és un sistema de suport a la decisió per ajudar els experts mèdics en el seu diagnòstic. L'arquitectura general permet la integració de diverses fonts de dades que corresponen als diferents perfils de metges i tècniques involucrades en un diagnòstic de melanoma. Cada font de dades s'utilitza per configurar un únic subsistema CBR que realitza una classificació específica. La combinació dels diferents subsistemes de CBR a través del protocol mèdic és el que dóna el diagnòstic final.

Seguint l'objectiu de l'ajuda al diagnòstic s'ha dissenyat i implementat DERMA com a resposta als diferents problemes i reptes plantejats pels experts del domini. Cada meta que hem abordat ha estat coberta per un objectiu del sistema. Com hem vist, les diferents fites assolides per DERMA han aconseguit millorar els resultats dels passos anteriors. El treball es va començar amb una sola classificació, utilitzant un únic sistema CBR amb un sol tipus de dades. El primer pas és el que va permetre determinar la precisió de base. Més tard es va proposar un diagnòstic col·laboratiu seguint el protocol mèdic, en què els diferents subsistemes fan un pronòstic unificat amb diferents fonts de dades. Aquest mòdul du a terme el mateix procés utilitzat pels experts. En el procediment mèdic els diagnòstics es generen a partir de la combinació del criteri dels diferents experts en reunions de diagnòstic. La col·laboració bàsica es va perfeccionar amb l'aplicació de millores en l'organització de la CM i el procés de col·laboració. Les optimitzacions del sistema es van fer per millorar la sensibilitat i especificitat que demanaven els experts. Finalment, el treball es va estendre per a l'ús de dades de múltiples etiquetes seguint les característiques de domini. L'últim pas ha ofert bons resultats i manté la porta oberta a conjunts de dades més rics. A nivell de resultats podem dir que DERMA és una eina adient per a l'ajuda al diagnòstic de càncer de melanoma. Cal, a més, destacar que els resultats han estat validats i donats per positius pels experts del domini. La validació dels resultats per part dels experts mèdics ha permès dir que s'ha assolit completament l'objectiu fixat a l'apartat 1.3.

Des d'un punt de vista més qualitatiu es vol remarcar que l'objectiu final de la tesi no ha estat només una aportació tècnica, sinó que es volia influir des d'un punt de vista més social. D'aquesta manera no només s'ha volgut avaluar la bondat dels mètodes aplicats en termes numèrics, sinó també en beneficis mèdics. El criteri mèdic per al diagnòstic és molt acurat, i permet una molt bona classificació dels pacients. Per això, una eina d'ajuda a aquest procés ha d'anar més enllà d'una classificació estàndard: ha de servir per ajudar a discernir en casos límit en els quals els mètodes tradicionals puguin oferir dubtes. I d'aquí neix la necessitat d'utilitzar tècniques que complementin l'algorisme diagnòstic emprat pels experts. La classificació multietiqueta, la millora de la memòria de casos o l'aplicació de regles de preprocessament han permès ajustar la classificació dels casos dubtosos permetent una classificació més fina. Atenent a aquests

condicionants es vol destacar que DERMA no només ofereix uns resultats competents, com hem vist a l'experimentació, sinó que, a més, són interessants a ulls del criteri mèdic. El sistema no pretén substituir cap expert sinó que busca ajudar-lo en les seves tasques i, per tant, pot utilitzar-se com una eina de suport als experts. DERMA vol col·laborar en la precocitat del diagnòstic. Aquest ha estat l'autèntic objectiu.

6.6 Línies de futur

L'objectiu principal de la tesi ha estat la creació d'una eina d'ajuda al diagnòstic de càncer de melanoma, propòsit que hem anat desenvolupant al llarg dels diferents apartats. Per crear l'eina hem comptat amb el suport de l'HCPB amb el qual hem analitzat les característiques que havia de tenir l'aplicació i hem resolt els diferents reptes que se'ns han anat plantejant durant la preparació de la tesi. El resultat ha estat el disseny, implementació i comprovació de DERMA, un sistema que dóna resposta a les peticions dels experts del domini (el càncer de melanoma) mitjançant una plataforma de suport a la decisió. Amb tot, encara queden portes obertes a noves aportacions tant dins de l'àmbit del càncer de melanoma com en altres àrees de coneixement. Les línies de treball de futur es divideixen en dues parts: (1) una primera que continua en l'àmbit del melanoma i que busca aprofundir en l'ajuda al diagnòstic a partir de nous conjunts de dades, i una altra (2) que fa el pas d'adaptar les tècniques utilitzades a DERMA per aplicar-les a altres dominis. A continuació veurem en detall les tasques futures.

Com hem vist al capítol 2, avui en dia la dificultat més destacada en les eines d'ajuda al diagnòstic de melanoma és la manca de conjunts de dades exhaustius. Els experts tracten amb un gran nombre de pacients però no tots els seus registres estan informatitzats i disponibles per ser utilitzats en una eina com DERMA. Actualment, des del GRSI s'està treballant amb la XCM per completar un conjunt de dades més ampli desenvolupant una aplicació que les gestioni. Quan es disposi d'aquest conjunt de dades amb totes les característiques del melanoma podrem tornar a testejar DERMA amb la nova informació. L'augment de dades no ha d'afectar el disseny de l'aplicació ja que està preparada per treballar amb tants mòduls (experts mèdics) com sigui necessari. A més, els nous estudis permetran la interacció de més tècniques diagnòstiques i experts del domini. D'aquesta manera es podrà afinar el resultat de l'eina d'ajuda al diagnòstic fins poder utilitzar-la en l'assistència mèdica habitual. Aquest és el principal objectiu de futur de la tesi.

En afegit, els sistemes híbrids amb raonament analògic, com la col·laboració multietiqueta amb CBR, són temes d'investigació d'actualitat i especialment interessants en àrees com les xarxes socials o el màrqueting perquè en aquests àmbits és fonamental comptar amb un marc flexible com el que ofereixen aquests mètodes. A més, són tècniques considerades altament fiables. Actualment, ja s'estan explotant alguns dels processos en aquests nous camps de manera que una segona línia de futur és traslladar les nostres propostes a noves àrees per fer-se càrrec de problemes d'altres àmbits. Una part del procés ja s'ha iniciat en l'àmbit multietiqueta utilitzant el sistema amb dades de fora del domini mèdic.

Bibliografia

A. Aamodt and E. Plaza. Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. *AI Communications*, 7:39–59, 1994. ISSN 0921-7126. 1

E. Armengol. Classification of melanomas in situ using knowledge discovery with explained case-based reasoning. *Artificial Intelligence in Medicine*, 51(2):93 – 105, 2011. ISSN 0933-3657. doi: http://dx.doi.org/10.1016/j.artmed.2010.09.001. 4

E. Armengol and S. Puig. Combining two lazy learning methods for classification and knowledge discovery. Senart, Paris, 2011. INSTICC, INSTICC. 4

J. Avila, E. Gibaja, and S. Ventura. *Multi-label Classification with Gene Expression Programming*, volume 5572 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 2009. 33

N. Cruz-Ramírez, E. Mezura-Montes, M. Y. Ameca-Alducin, E. Martín-Del-Campo-Mena, H. G. Acosta-Mesa, P.-C. Nancy, A. Guerra-Hernández, G. d. J. Hoyos-Rivera, and R. E. Barrientos-Martínez. Evaluation of the diagnostic power of thermography in breast cancer using bayesian network classifiers. *Computational and Mathematical Methods in Medicine*, 2013. doi: http://dx.doi.org/10.1155/2013/264246. 4

T. G. Dietterich. Ensemble learning. In *The Handbook of Brain Theory and Neural Networks*, pages 405–408. The MIT Press, 2002. 3

A. Fornells, E. Armengol, E. Golobardes, S. Puig, and J. Malvehy. Experiences using clustering and generalizations for knowledge discovery in melanomas domain. In *Advances in Data Mining. Medical Applications,E-Commerce, Marketing, and Theoretical Aspects*, volume 5077, pages 57–71. Springer, 2008a. ISBN 978-3-540-70717-2. 4, 12

A. Fornells, E. Golobardes, J. Martorell, and J. Garrell. Patterns out of cases using kohonen maps in breast cancer diagnosis. *International Journal of Neural Systems*, 18:33–43, 2008b. 4

J. Han and M. Kamber. Data mining: Concepts and techniques, 2006. 4, 6

A. Jain, A. Jain, and S. Jain. *Artificial Intelligence Techniques in Breast Cancer Diagnosis and Prognosis*. Series in Machine Perception and Artificial Intelligence. World Scientific Publishing Company, 2000. ISBN 9789810243746. 4

F. V. Jensen and T. D. Nielsen. *Bayesian networks and decision graphs*. Information science and statistics. Springer, New York, 2007. ISBN 0-387-68281-3. 4

K. D. Jong. *Evolutionary Computation: A Unified Approach*. The MIT Press, 2006. ISBN 0262041944. 1

J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. Vol. 1.*, pages 281–297, 1967. 12

O. Maimon and L. Rokach. *Data Mining and Knowledge Discovery Handbook*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005. ISBN 0387244352, 9780387244358. 1

W. S. McCulloch and W. Pitts. Neurocomputing: Foundations of research. chapter A Logical Calculus of the Ideas Immanent in Nervous Activity, pages 15–27. MIT Press, 1988. ISBN 0-262-01097-6. 2

J. E. McWhirter and L. Hoffman-Goetz. Visual images for patient skin self-examination and melanoma detection: A systematic review of published studies. *Journal of the American Academy of Dermatology*, 69:47–55, 2013. 3

P. Melville and R. J. Mooney. Diverse ensembles for active learning. In *ICML 04: Twenty-first international conference on Machine learning*, 2004. 6, 18

R. Nicolas, E. Golobardes, A. Fornells, S. Puig, C. Carrera, and J. Malvehy. Identification of relevant knowledge for characterizing the melanoma domain. In *Advances in Soft Computing*, volume 49 2009, pages 55–59. Springer Berlin-Heidelberg, 2008a. ISBN 978-3-540-85860-7. 5, 10, 11, 12, 18

R. Nicolas, E. Golobardes, A. Fornells, S. Puig, and J. Malvehy. Estudi de les caracteristiques del domini del melanoma (code: 071208). Technical report, EALS-URL, 2008b. 12

R. Nicolas, E. Golobardes, A. Fornells, S. Segura, S. Puig, C. Carrera, J. Palou, and J. Malvehy. Using ensemble-based reasoning to help experts in melanoma diagnosis. In *Frontiers in Artificial Intelligence and Applications*, volume 184, pages 178–185. IOS Press, 2008c. ISBN 978-1-58603-925-7. 18, 19, 21, 30

R. Nicolas, D. Vernet, E. Golobardes, A. Fornells, F. de la Torre, and S. Puig. Distance metric learning in a collaborative melanoma diagnosis system with case-based reasoning. In *Proceedings of the 14th United Kingdom Workshop on Case-Based Reasoning at The 29th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence, CMS Press, University of Greenwich, UK*, pages 58–66, 2009a. 14, 15

R. Nicolas, D. Vernet, E. Golobardes, A. Fornells, S. Puig, and J. Malvehy. Improving the combination of cbr systems with preprocessing rules in melanoma domain. In

Workshop Proceedings of the 8th International Conference on Case-Based Reasoning, Seattle, WA, USA, pages 225–234, 2009b. 18, 20, 21

R. Nicolas, A. Fornells, E. Golobardes, G. Corral, S. Puig, and J. Malvehy. Melanoma diagnosis based on collaborative multi-label reasoning. In *Frontiers in Artificial Intelligence and Applications*, volume 256, pages 283 – 292. IOS Press, 2013a. doi: 10.3233/978-1-61499-320-9-283. 25, 26

R. Nicolas, A. Sancho-Asensio, E. Golobardes, A. Fornells, and A. Orriols-Puig. Multi-label classification based on analog reasoning. *Expert Systems With Applications Journal*, 40:5924–5931, 2013b. doi: http://dx.doi.org/10.1016/j.eswa.2013.05.004. 6, 24, 25, 26, 32, 33, 34

R. Nicolas, A. Fornells, E. Golobardes, G. Corral, S. Puig, and J. Malvehy. Derma: A melanoma diagnosis platform based on collaborative multi-label analog reasoning. *The Scientific World Journal*, 2014:11 pages, 2014. doi: http://dx.doi.org/10.1155/2014/351518. 6, 26

S. Ontañón. *Ensemble Case Based Learning for Multi-Agent Systems*. VDM Verlag, 2008. ISBN 3836474301. 18

S. J. Russell, P. Norvig, J. F. Candy, J. M. Malik, and D. D. Edwards. *Artificial Intelligence: A Modern Approach*. Prentice-Hall, Inc., 1996. ISBN 0-13-103805-2. 1

N. Singh, A. G. Mohapatra, and G. Kanungo. Breast cancer mass detection in mammograms using k-means and fuzzy c-means clustering. *International Journal of Computer Applications*, 22(2):15–21, 2011. 4

D. Vernet and E. Golobardes. Adding Unsupervised Learning in the Case-Based Organization. In *Frontiers in Artificial Intelligence and Applications*, volume 100, pages 307–315. IOS Press, 2003. 12

D. Vernet, R. Nicolas, E. Golobardes, A. Fornells, C. Garriga, S. Puig, and J. Malvehy. Pattern discovery in melanoma domain using partitional clustering. In *Frontiers in Artificial Intelligence and Applications*, volume 184, pages 323–330. IOS Press, 2008. ISBN 978-1-58603-925-7. 12, 28

D. Vernet, R. Nicolas, E. Golobardes, A. Fornells, and A. Garcia-Piquer. Intelligent tutoring system framework for the acquisition of knowledge and competences. In *40th ASEE/IEEE Frontiers in Education (FIE) Conference*, pages 111–112. IEEE, 2010. ISBN 978-1-4244-6259-9. 26

M.-L. Zhang and Z.-H. Zhou. A K-Nearest Neighbor based Algorithm for Multi-Label Classification. In *Granular Computing*, pages 718–721. IEEE, 2005. ISBN 0-7803-9017-2. 6, 24

"El futur té molts noms. Per als dèbils, allò que no es pot aconseguir. Per als temeraris, allò desconegut. Per als valents, l'oportunitat"

Víctor Hugo

A Acrònims

ABCD Asimetria, Contorn, Color, Diàmetre / *Asymmetry, Border, Color, Diameter*

AGAUR Agència de Gestió d'Ajuts Universitaris i de Recerca

AI Intel·ligència Artificial / *Artificial Intelligence*

AAD Acadèmia Americana de Dermatologia / *American Academy Dermatology*

AAM Mètodes d'Adaptació d'Algorismes / *Algorithm Adaptation Methods*

CBR Raonament Basat en Casos / *Case-Based Reasoning*

CM Memòria de Casos / *Case Memory*

CMU *Carnegie Mellon University*

COMEDI-CBR *COllaborative MElanoma DIagnosis using CBR*

DERMA *Melanoma Diagnosis based on Collaborative Multi-Label Reasoning*

DM Mineria de Dades / *Data Mining*

DML Aprenentatge de Mètriques de Distància / *Distance Metric Learning*

EC Computació Evolutiva / *Evolutionary Computation*

EL Aprenentatge Conjunt / *Ensemble Learning*

EPO Ens Promotor Observador

GRSI Grup de Recerca en Sistemes Intel·ligents

HCPB Hospital Clínic i Provincial de Barcelona

IDIBAPS Institut d'Investigacions Biomèdiques August Pi i Sunyer

KNN K-Veí més Proper / *K-Nearest Neighbor*

MEDIBE *MElanoma DIagnosis Based on Ensembles*

MID-CBR *Marco Integrador para el Desarrollo de Sistemas de CBR*

MlKnn *Multi-Label K-nearest neighbor*

MlCBR *Multi-Label Case-Based Reasoning*

ME Multietiqueta

NN Xarxes Neuronals / *Neural Networks*

PR Reutilització Probabilística / *Probabilistic Reuse*

PRE Reutilització Probabilística amb Experiència / *Probabilistic Reuse based on Experience*

RAkEL *RAndom k-labELsets*

RB-MEDIBE *Rule Based MElanoma DIagnosis Based on Ensembles*

ULIC *Unsupervised Learning In CBR*

XCM Xarxa Catalana de Melanoma

"No hi ha exèrcit que pugui resistir la força d'una idea a la que li ha arribat l'hora"

Winston Churchill

B

Publicacions

DERMA: A melanoma diagnosis platform based on collaborative multi-label analog reasoning. *The Scientific World Journal*, 2014

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2014, Article ID 351518, 11 pages
http://dx.doi.org/10.1155/2014/351518

Research Article

DERMA: A Melanoma Diagnosis Platform Based on Collaborative Multilabel Analog Reasoning

Ruben Nicolas,[1] Albert Fornells,[1] Elisabet Golobardes,[1] Guiomar Corral,[1] Susana Puig,[2] and Josep Malvehy[2]

[1] *La Salle, Ramon Llull University, Quatre Camins 2, 08022 Barcelona, Spain*
[2] *Melanoma Unit, Dermatology Department, Clinic Hospital, Rosselló 149, 08036 Barcelona, Spain*

Correspondence should be addressed to Ruben Nicolas; rnicolas@salle.url.edu

Received 7 August 2013; Accepted 12 November 2013; Published 21 January 2014

Academic Editors: S. Jagannathan and S. Sessa

The number of melanoma cancer-related death has increased over the last few years due to the new solar habits. Early diagnosis has become the best prevention method. This work presents a melanoma diagnosis architecture based on the collaboration of several multilabel case-based reasoning subsystems called DERMA. The system has to face up several challenges that include data characterization, pattern matching, reliable diagnosis, and self-explanation capabilities. Experiments using subsystems specialized in confocal and dermoscopy images have provided promising results for helping experts to assess melanoma diagnosis.

1. Introduction

Melanoma is growing in importance because it is increasingly more prevalent in our society and it affects people of any age. Although it is not the most common skin cancer, if it is not early treated, its mortality is around twenty percent, according to the American Academy of Dermatology [1]. The most important difficulties related to early diagnosis are that it is a problem with a non trivial classification process, given the high volume of data, different experts, and types of diagnosis and the fact that there are not enough clear classification patterns. The characteristics of the problem have fomented the application of artificial intelligence techniques to exploit data and help experts in the early diagnosis. This work describes DERMA, a melanoma diagnosis architecture created as a result of the collaboration between the department of dermatology at the Hospital Clinic of Barcelona (HCPB) and the Institute of Biomedical Research August Pi i Sunyer (IDIBAPS). DERMA is a collaborative architecture among several subsystems specialized in different kinds of data sources. More specifically, the current version is based on the collaboration of two multilabel case-based reasoning (CBR) [2] systems that use confocal and dermoscopy images, respectively, which are the most important image analysis in melanoma cancer to date [3, 4]. CBR is used as an engine due to its self-explanation capabilities extracted from solving new problems from past experiences, which are important for experts to understand the results. Thus, on the other hand, the multilabel mode [5] means that a new case is classified in several subclasses, so this additional information helps experts to understand the melanoma case more accurately because complex patterns are better described as a set of simpler patterns. The collaboration of CBR systems enables the replication of the medical protocols from the different expert profiles. Moreover, the CBR data is reorganized using distance metric learning [6] to promote the separation between malignant and nonmalignant cases. Therefore, distance metric learning and the multilabel collaboration scheme result in an improvement in the sensitivity and specificity of the diagnosis, which is precisely what medical experts are looking for.

The following sections are described as follows. Section 2 summarizes some related work to contextualize the proposal. Section 3 presents the DERMA architecture and each one

of its modules. Next, Section 4 highlights the main results obtained. Finally, Section 5 ends with the conclusions and further work.

2. Related work

Artificial intelligence (AI) techniques have been used with outstanding results in different knowledge areas such as marketing, education, or medicine. This is because these kinds of problems have so much information which is not directly processable by the human mind. To overcome this difficulty we need techniques to extract patterns and to deal with this information. Nowadays research in artificial intelligence for cancer is an outstanding research topic supported by different institutions and contests such as the Google Science Fair which awarded a breast cancer research project in 2012 or the Intel International Science and Engineering Fair 2013 that won a leukemia project.

The range of artificial intelligence techniques applied in medical problems is large and covers four groups of data mining methods evolved in these processes: clustering, association rules, classification, and regression. Regardless of the outstanding groups which are clustering, the goal is to decompose the problem and try to model it in a proper manner. Therefore, association rules, which focus the interest on why things happen, are of great relevance. In some cases, clustering and association rules are used as base methods to apply classification and regression processes. Considering these characteristics we could break the problem into two frequent groups: methods to find patterns and tools to aid decision making (recommender systems) with some examples of works on these areas.

The objective of aiding decision making has been addressed in several types of cancer such as breast cancer and has been treated from different points of view [7]. One goal is to achieve an automatic feature extraction from breast images in order to detail their characteristics. The work performed by [8] meets this objective through the use of k-means and fuzzy c-means clustering techniques. With the aim of avoiding invasive techniques with psychological, health, and economical consequences, some works that use techniques such as thermography for diagnosis appear. An approach using Bayesian networks is presented by [9] in order to fit this target. DESMAI framework [10] allows experts in breast cancer to explore digital mammography databases according to a certain topology criteria when they need to decide whether a sample is benign or malignant. This work was performed through a variant of a case-based reasoning system featured by organizing the case memory using self-organizing maps (SOM). In melanoma cancer, one approach is to construct a domain model that combines learning methods with different characteristics [11].

Considering the recommendation methods group we found [12] which proposes an automatic way to build decision support systems by means of combining several machine learning techniques using a metalearning approach based on grammar evolution. In the particular case of melanoma cancer there are works that enable knowledge discovery such as [13] that uses SOM to identify groups of similar melanoma and creates descriptions of clusters that are used as explanations for experts. To support dermatologists in assessing the classification of skin lesions using dermoscopy in order to asses prior to extraction, [14] uses a combination of case-based reasoning and clustering for generating a domain theory to classify melanomas in situ. In addition there are works that try to automatize concrete parts of the diagnosis, such as dermoscopy analysis, using artificial intelligence [15, 16].

After the study of melanoma cancer problem we found that we are not just interested in a correct data analysis but we seek, above all, reliability. So our challenges properly represent knowledge, analyze each type of information, and establish collaborations between experts and tests. Thus we can not directly use any of the methods used until now, so we need a proposal specific to the problem that merges data mining techniques with the medical protocol. Attending to all these considerations DERMA is an interesting proposal for medical experts due to its ad hoc adaptation to the problem.

3. DERMA: Melanoma Diagnosis Based on Collaborative Multilabel Analog Reasoning

DERMA is a platform that aids medical experts in melanoma diagnosis. Figure 1 describes the DERMA architecture which addresses four different challenges identified during the collaboration with HCPB that range from data acquisition to diagnosis. The first challenge focuses on the creation of a melanoma ontology [17] based on a characterization of the domain performed through interviews with melanoma experts and the study of melanoma patterns. The second one is to create specialized subsystems in order to work with the different data sources. CBR is selected due to its suitability for working in environments where self-explanations are required. This step also considers how to organize the knowledge bases in a proper manner in order to improve its performance through distance metric learning [6]. The third challenge is to define a collaborative scheme [18] between the independent CBR subsystems based on the way by which experts work and which also includes mechanisms to manage exceptional situations. Finally, the last challenge is related to the complex task of making classifications of nontrivial patterns. In this sense, DERMA may be able to learn and diagnose better if richer patterns could be represented as a set of simple patterns. For this reason, we decide to extend the single-label CBR subsystems to multilabel [5] CBR subsystems. The next subsections describe these challenges in detail.

3.1. Challenge 1: Melanoma Characterization. The first step in the development of any knowledge-based system is to identify, understand, and gather data associated with the problem. These steps are nontrivial when the system is related to the health sciences domain because these data are usually characterized as being heterogenous and coming from several medical profiles involved in the prognosis. Moreover, experts often label data according to their interests and background,

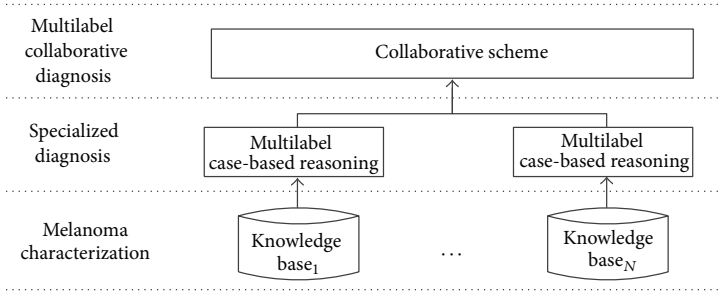

Figure 1: Melanoma diagnosis architecture based on collaborative multilabel reasoning.

so attributes with different names may have the same meaning. Thus, the characterization and understanding of the relationships between all the data sources for planning the gathering of knowledge is a nontrivial task that requires time and consensus between experts.

The concrete domain characterization was performed using data from more than three thousand patients with melanoma and contained reports from dermatologists, oncologists, surgeons, pathologists, and other specialists working in HCPB. As in the great majority of medical problems, data was heterogenous and distributed in different plain databases and many attributes were represented and stored differently according to the expert. For this reason, an ontology [17] with more than forty concepts was defined using the experts' point of view and data from international studies that examine specific aspects of the domain [19] divided in five groups: (1) person and family, (2) generic medical information, (3) tumors, (4) metastasis, and (5) controls and studies. Although this unified point of view permitted the integration between all data sources as Figure 2 shows through a relation model, there were pieces of information that could not be integrated and used in platform tests because experts did not have records regarding all patients. This is the reason why we decided to focus our work on the usage of nevus images analysis due to its availability and being outstanding between other data. More specifically, dermoscopy and confocal images were selected because they are two of the most promising techniques of image analysis for the diagnosis of melanoma. Dermoscopy is based on a microscopic image created by epiluminiscence microscopy (x10.30) and confocal reflectance is generated by the reflection of a coherent laser (x100) resolution at the level of the cell [3, 4].

In order to test if it was possible to identify equivalent patterns using the same data used by medical experts, we applied several data mining techniques. Melanoma diagnosis is based mainly on the ABCD rule which considers the following characteristics that are typically observed in this type of tumor: (A) a diameter greater than 5 mm, (B) color variation, (C) asymmetry, and (D) jagged edges. We tested K-means [20] and SOM [21] for extracting patterns due to our previous experiences in breast cancer diagnosis using these techniques [10]. K-means algorithm makes a partition of the domain in K clusters and SOM translates complex and nonlinear statistical relations contained in high-dimensional data into simple geometric relations on a low-dimensional space which provide an optimal organization. The results provided equivalent patterns to the ones identified by ABCD rules [13, 22] attending to medical criteria.

3.2. Challenge 2: Specific Diagnosis Using the Most Useful Knowledge. CBR systems solve new problems through an analogical procedure based on experiences represented by a set of cases stored in a case memory. Thus, CBR is able to justify the obtained solutions using analogies with previous problems which is crucial for experts. The way in which CBR works can be summarized in the following steps: (1) it retrieves the most similar cases from the case memory with a similarity function; (2) it adapts them to propose a new solution; (3) it checks if this solution is valid; and finally (4) it retains the useful information of the prognostic if it is necessary. All CBR steps turn around the case memory and its organization and how cases are retrieved determine its performance in terms of accuracy, specificity, sensitivity, and computational time [23].

There are two main possible memory organizations: flat and structured. A flat organization is the simplest way because cases are stored sequentially in a list. In such situations, the strategy to classify a new problem is to sequentially compare it with all the cases in that list using some similarity measure. The main shortcomings of this approach are that the more cases the case base contains, the higher the time of retrieval is and that the lack of organization may imply that useful cases are skipped. Structured memory organization focuses on improving both issues and many authors have tackled this issue from many points of view, such as representing the attributes in tree structures [24] or graphs [25], grouping cases by their similarity [26], and applying knowledge-intensive approaches [27] or data-intensive approaches [28]. Independent of the case memory organization, a distance function needs to be defined for comparing the similarity of cases. The ideal similarity function definition is not a trivial task because it depends on the domain and how data is related. There are even works that try to discover this similarity function using algorithms based on genetic algorithms [29]; the application of standard similarity functions such as Euclidean distance is the most frequent solution due to the complexity of identifying reliable distance metrics.

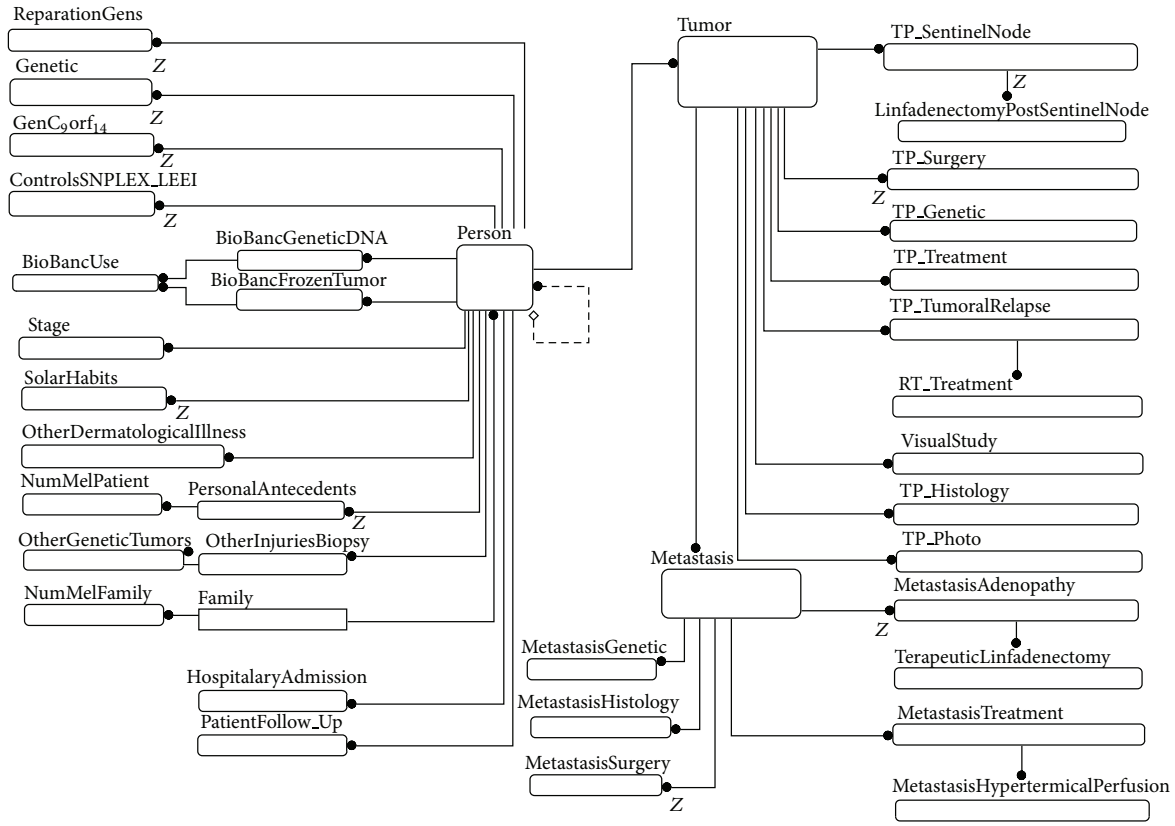

Figure 2: The melanoma relational model permits the definition of how to integrate data gathered from the different medical profiles. The model considers patient data, their family, generic information, tumors, metastasis, controls, and studies.

Owing to the importance of better determination of positive and negative melanoma cases we propose taking case memory organization and similarity function definition through a data organization based on distance metric learning (DML) [6]. DML is a technique used to identify a suitable distance metric based on the data projection that can be divided into four families [30]. The first two families are based on the supervision of the method: supervised and unsupervised DML. The last two families are based on a more concrete classification: based on support vector machines or kernel methods. In our case and in response to the characteristics of the problem, we are working with the supervised family. With this method we learn a metric that keeps all the data points from the same class close together and, at the same time, separates as far as possible the data points from different classes. We have learned a global distance metric that minimizes the distance between pairs of data included in the equivalence constraints and data pairs from the inequivalence constraints. With this process we obtained a case memory organized in a way that enables a better retrieval because positive and negative cases become distanced [31].

3.3. Challenge 3: A Global Diagnosis Using Independent and Specific Diagnosis. The way in which melanoma is diagnosed takes into account different data sources and this makes it easy to apply a collaborative approach to classify new patients. Figure 3 describes the medical process that experts consider, that is, effectively a collaborative process that determines if the new case is melanoma, basal cell carcinoma (BCC), or a nonmalignant tumor (melanocytic or not) according to the partial diagnosis using the confocal and dermoscopy images of the new patient.

The combination of approaches can be summarized [32] in (1) bagging, (2) boosting, and (3) stacking. Bagging and Boosting are based on the combination of the outputs using votes. In concrete bagging replicates N systems of the same approach but uses different data sources. In opposition boosting follows the same idea but defines models in order to complement them. Finally stacking [11] is based on heuristics that combine the outputs of several approaches. The most common voting methods [33] are (1) plurality, (2) contra-plurality, (3) borda-count, and (4) plurality with delete. All of them are based on the number of votes of a class (plurality) but with differences in the addition of plurality and decision of better class.

There are several works that use collaborative systems that permit an improvement in well-known algorithms such as clustering using collaboration [34], to allow the classification using data of different complexity [35], or with different

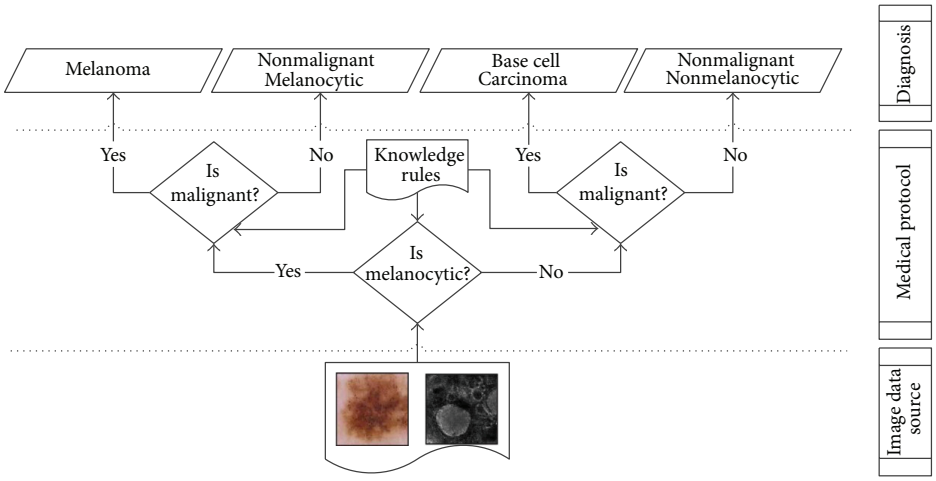

Figure 3: Medical diagnosis protocol schema followed by dermatological cancer experts.

types of medical information [12]. There are so many general collaborative systems [36], but in melanoma classification we must consider specific characteristics that need the use of medical knowledge. The collaboration protocol should follow the one used by experts in melanoma that is to combine different decisions from different systems to build a more reliable solution using the individual ones, as it has been done in other problems [37]. In our case we are working with two different points of view (confocal and dermoscopic) from which we select the best classification from one of the systems depending on different criteria [38]. DERMA functional schema is shown in Figure 4 where we define two specialized CBR modules that follow the medical protocol for classification and later we combine the obtained results through collaborative criteria. Thus, the concrete characteristics of the domain [19] make it necessary to employ a different method from the general one. As we are using different attributes of the same data in each system, then the independence of the data is guaranteed, in contrast to the standard bagging. Analyzing the classification attributes, the voting method should be based on plurality, albeit with some specific conditions requested by medical researchers, who place more importance to the information from confocal microscopy because they consider it to be more reliable.

On the other hand, bearing in mind that the main goals are the improvement of the classification and minimizing the false negative situations, a knowledge rules module is introduced to ensure reliability in exceptional situations due to data oddities in a second layer of the collaborative system [39]. This module preprocesses the input data and creates a set of rules to help the whole classifier. It has been done using clustering in order to discover new patterns on the medical domain [22] and to detect particular behaviors on the data. Despite using a similar idea of [40], we preprocess the data in a nonbased interval way, where concrete values are detected and encapsulated in a rule. Moreover, our rules do not depend on each other and attributes are analyzed independently. The idea is to weight the single classification of each subsystem according to the reliability of the retrieved cases. The reliability of a case is based on a set of rules previously extracted from data. This new step adds a fine tuning to the classifier collaboration that leads to an improvement of the final classification.

3.4. Challenge 4: Multilabel Diagnosis. During the integration of the medical protocol an interesting aspect shows up: the final diagnosis is obtained from considering different medical profiles and/or classification patterns. This is the principle of multilabel classification problems where there is not a single class, but rather elements that are carved in parts to avoid information loss. We consider that the problem is better represented as two nondisjunctive classes such as melanocytic and malignant ones instead of using just a single class.

In the last few years we have witnessed an increase in the use of multilabel systems due to their better fitting to real problems and their ability to avoid information loss. Existing works to date have been divided into two distinct families. The first option is to adapt the dataset to work with single label algorithms instead of designing new algorithms. This group of techniques are known as problem transformation methods (PTM) and the main problem is that the unification of the different labels in a unique label is the loss of information. The most competent works in PTM for multilabel classification are MlKnn and RAkEL [41, 42] and both are recognized by the community as reference algorithms. MlKnn is a theoretical approach to multilabel classification which adapts the k recovered cases from the classical k nearest neighbor algorithm (Knn) [43] to multiple label problems. RAkEL is an ensemble platform that permits the classification of multilabel datasets by dealing with each label separately and combining the single-label results. It is publicly available through WEKA [44]. There are other interesting works in this field used by the community such as [45] where the authors present a pruned transformation

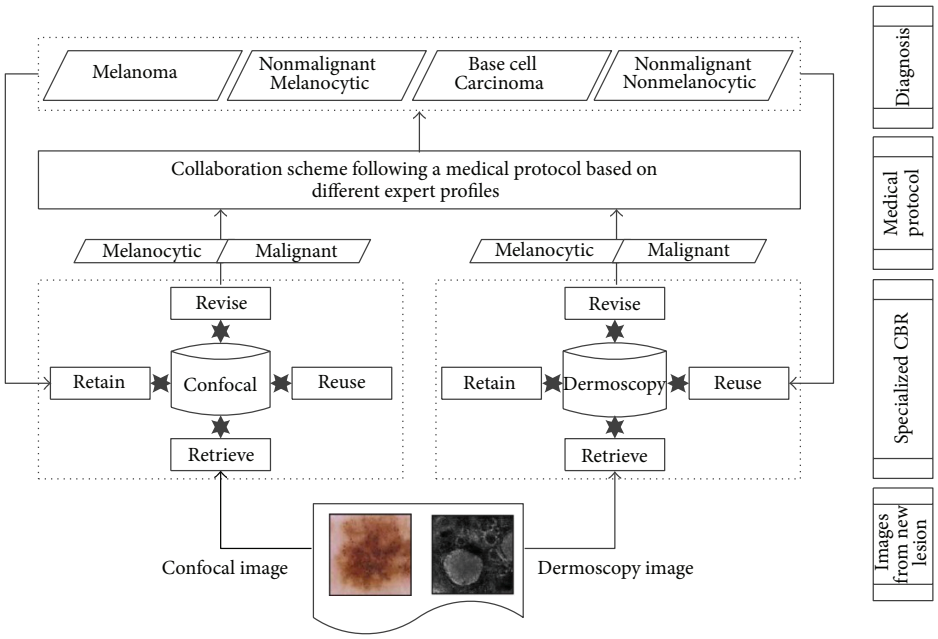

Figure 4: DERMA is based on a collaborative scheme between specialized CBR subsystems for melanoma cancer diagnosis following the medical diagnosis protocol.

that combines key points of several previous approaches and [46] that uses neural networks for multilabel classification. The second family addresses the problem with a modification of the classical algorithms to work purely multilabel and is known as algorithm adaptation methods (AAM). The most influential works in AAM include a boosting algorithm for text categorization [47], an adaptation of C4.5 [48] to deal with multilabel biological data [49], and a system that combines ranking methods with a predictor of the sets size [50].

Because we do not want to lose any information during the classification process, we focused on AAM approach. More specifically, we extended the phases of all CBR subsystems for working in multilabel mode. The four stages are designed as follows: (1) retrieval step follows the same pattern as the regular CBR algorithm. The system chooses the most similar cases according to the value obtained from a similarity function. The cases of the case memory that are more similar to the new case are the retrieved ones; (2) reusal stage has been adapted to multilabel classification in a probabilistic manner. It is based on the idea of counting the occurrences of each label and considers it positive if more than a half of the k retrieved cases have this positive label. This is similar to the idea proposed by multilabel k-nearest neighbor algorithm. A second step on the reuse phase considers the experience of each retrieved case (obtained through the retain feedback) weighting the recovered cases in an appropriate manner; (3) retaining phase to keep information on successes of the retrieved cases. This is the information by which we will weight the retrieved cases considering the successes of previous classifications; (4) revision, as in single-label, is performed by a medical expert.

4. Experiments and Results

The active collaboration with medical experts from HCPB has played an important role in all the steps taken to design DERMA and also in the preparation and study of experiments. Challenge 1 allowed us to characterize the domain taking into account the different medical profiles in order to define the ideal structure of the dataset. We are currently working with the Catalan Network of Melanoma in order to complete this dataset. Given the data available in melanoma database we have done two kinds of experiments: the first block covers the usage of dermoscopy and confocal data for testing the collaborative approach of the CBR subsystems in both data, and the second one works with multilabel melanoma data in order to test the complete DERMA.

The next points summarize the data used for experimentation and the most important milestones achieved and the outstanding results following the cited two blocks of experiments.

4.1. Testbed. The most used techniques to gather information from tissue are the dermoscopic and the confocal analysis. Confocal microscope is the most precise and the one that medical experts consider as world class. Nevertheless, a negative point is that the confocal analysis is a long and expensive test, so the number of available cases is limited. Due to this situation, the data set used in this work is composed of 150 cases of suspicious lesions. For all these cases we have information related to confocal and dermatoscopic images and the histology that corroborated diagnosis. Attending to the considerations of the medical experts that have created this set, it includes enough cases from each kind of illness

to be representative of the domain. Then, in medical terms it is an appropriated case memory for this study. Detailing the instances, dermoscopy information has forty-one fields and confocal microscopy, due to its higher resolution, contributes to data from eighty-three different attributes. This data has been configured in two different manners: the first as a single-label set of dataset where each classification process uses the appropriate attributes to classify one class and the second exploits all the properties of the data to classify all the possible classes at a time.

4.2. Experimental Framework and Configurations. Experimentation has been carried out according to the medical purposes in order to analyze the most interesting results for experts, such as false positives. The experimentation with medical data from confocal and dermoscopy images has tested DERMA using different configurations and analyzing sensitivity, specificity, and accuracy. The study considers two independent CBR subsystems with a basic decision combination, with the use of rules obtained through the use of preprocessing algorithms and with the application of DML to the original data. In addition, we tested the accuracy of the two independent CBR systems (one for confocal data and another for dermatoscopy). All the CBR systems used in experimentation are configured with one-nearest neighbor algorithm with normalized Euclidean distance as retrieve function and classify a single class at a time. In the case of the plain combination platform, the medical consensus is to use 0.5 as confocal threshold and double of the distance between the new case and the best confocal case as dermatological one. And the other stages use the threshold weighted by the preprocessed rules. This experiment framework has been tested applying a leave one out to the original data to obtain the average accuracy of those systems. The final challenge was focused on the use of multilabel data and, as a consequence, the use of multilabel CBR subsystems in our collaborative platform. For this experimentation we considered two classes in each instance (melanocytic and malignant) that offer experts the whole range of classifications of a nevus as in the single-label case. In addition we have performed a t-test with 95% confidence level between each configuration of DERMA and the previous one to establish the results of significance.

4.3. Results Using Single-Label Data. Having described the experimental framework we were able to analyze the results obtained. Table 1 shows sensitivity, specificity, and accuracy rates classifying new injuries using the two independent CBR and the three different collaboration schemes defined in DERMA: with rules, without them, and with the DML module. The results on Table 1 show the four possible classes that are considered by the medical protocol. The statistical significance of the results is represented with an ↑ if it is significantly better and a (—) if it is equivalent. The results obtained highlight that (1) the different layers added to DERMA achieve better classification results than the plain combination DERMA or the independent noncollaborative systems; (2) the use of the combination of both types of images with the help of preprocessing obtained rules leads to an important increase in sensitivity and specificity rates, the most important results for medical experts; (3) using the DML technique in order to better classify the new cases, we accomplish the desire of medical experts that is to avoid false negatives allowing the successful diagnosis of all patients; (4) it is easier to classify nonmalignant cases, it seems to be related to the fact that we use a real world dataset that fits the characteristics of the population and there are more patients with nonmalignant cases than with malignant ones; (5) the nonmelanocytic cases are better classified due to the characteristics of the problem; (6) the t-test results show that any improvement of DERMA has a significantly negative effect.

4.4. Results Using Multilabel Data. The results obtained by DERMA are shown in Table 2 and show the percentage of successes considering the four possible nevus classifications. This table is formed by the sensitivity, specificity, and accuracy results obtained using a noncollaborative protocol with confocal and dermoscopy data, respectively, and with a collaborative pattern with the combination of both data using plain combination, preprocess rules, and distance metric learning. As in single label the significance is represented with an ↑ if the result is significantly better and a (—) if it is equivalent. The results obtained highlight that (1) as in single-label classification the different layers added to DERMA achieve better classification results than the plain combination DERMA or the independent noncollaborative systems; (2) the results using the whole system permit the successful diagnosis of all cases. Although we achieve hundred percent accuracy we are not facing a foolproof system but one which knows how to work with the peculiarities of the domain, just as medical experts; (3) the classification of all classes in an unique step does not lose any kind of information; (4) each enhancement of DERMA is positive with statistical significance or, at least, equivalent in comparison to the nonuse of the improvement.

In addition to this experimentation and in order to test the strength of the method we made a previous study where we tested the multilabel classification part of DERMA with other kinds of data (due to the absence of more melanoma multilabel datasets) and in comparison with other multilabel classification platforms [51]. This work allowed us to tune the characteristics of our platform and to validate it as a competent method in this kind of classification. As general purpose repositories such as UCI [52] do not give enough multilabel datasets, we tested DERMA with seven synthetic datasets and the three most common real world multilabel datasets [5]. The obtained results show that DERMA results are equivalent to the ones obtained by reference platforms RAkEL and MlKnn. We would like to highlight that the multilabel classification reduces the steps required to obtain the same result, thus reducing the computational costs.

4.5. Global Results and Discussion. Once the results of our platform have been presented in terms of performance, we will discuss the system: its characteristics and its applications. As we have seen DERMA is an on-demand application that

Table 1: Sensitivity, specificity, and accuracy results obtained in melanocytic, melanoma, and BCC classification through the different DERMA challenges: the noncollaborative CBR classification which only uses dermoscopy data, the noncollaborative CBR classification which only uses confocal data, the plain collaborative system, the collaborative system that enhances the collaboration with preprocessing rules, and the collaborative system with a DML organized case memory. Each result shows the t-test comparison between the result obtained on this DERMA configuration in comparison with the previous one using 95% of confidence level. This is presented with an (↑) if it is significantly better and (—) if there is no significant difference.

	Nonmalignant		Malignant	
	Melanocytic	Nonmelanocytic	Melanocytic (melanoma)	Nonmelanocytic (BCC)
		Sensitivity results		
Dermoscopy CBR	75%	80%	73%	81%
Confocal CBR	74% (—)	92% (↑)	73% (—)	92% (↑)
Collaborative	80% (↑)	94% (—)	70% (—)	92% (—)
Collaborative + rules	95% (↑)	95% (—)	81% (↑)	92% (—)
Collaborative + rules + DML	100% (↑)	100% (↑)	100% (↑)	100% (↑)
		Specificity results		
Dermoscopy CBR	95%	99%	92%	96%
Confocal CBR	99% (↑)	98% (—)	96% (↑)	95% (—)
Collaborative	96% (—)	97% (—)	95% (—)	96% (—)
Collaborative + rules	99% (—)	99% (—)	98% (↑)	100% (↑)
Collaborative + rules + DML	100% (—)	100% (—)	100% (↑)	100% (—)
		Accuracy results		
Dermoscopy CBR	90%	96%	87%	96%
Confocal CBR	88% (—)	95% (—)	90% (↑)	95% (—)
Collaborative	92% (↑)	94% (—)	89% (—)	95% (—)
Collaborative + rules	98% (↑)	99% (↑)	94% (↑)	99% (↑)
Collaborative + rules + DML	100% (↑)	100% (↑)	100% (↑)	100% (↑)

Table 2: Sensitivity, specificity, and accuracy results obtained in multilabel classification using dermoscopy data, confocal data, and both types of data with a collaborative system and using the different DERMA modules. Each result shows the t-test comparison between the result obtained on this DERMA configuration in comparison with the previous one using 95% of confidence level. This is presented with an (↑) if it is significantly better and (—) if there is no significant difference.

	Sensitivity	Specificity	Accuracy
Multilabel dermoscopy	86%	89%	92%
Multilabel confocal	93% (↑)	97% (↑)	96% (↑)
Multilabel collaborative	91% (—)	96% (—)	95% (—)
Multilabel collaborative + rules	94% (↑)	99% (↑)	98% (↑)
Multilabel collaborative + rules + DML	100% (↑)	100% (—)	100% (↑)

fits the needs of medical experts in melanoma diagnosis. Attending to the analysis of possible helpful and harmful issues from internal and external origins, we could point that DERMA has its strengths in the fact that it improves the sensitivity and specificity rates, which is important for experts, and gives classification explanations. These explanations are based on the retrieved cases that provide the medical experts with an explanation of the similarity between the new case and the prediction. Moreover it is based on an increasingly common cancer that, attending to the American Academy of Dermatology, improves the recovery results with a proper early diagnosis offering an important opportunity to research. On the other hand, we must deal with external threats such as data availability, the medial protocol that is dynamic, and the changes in data from new studies. All theses weaknesses are being covered through a data platform and with the scalability and fine tuning of DERMA that allows a wide range of changes.

5. Conclusions and Further Work

Melanoma cancer is a growing problem in our society due to the increasing number of cases. The characteristics of this disease and the different types of techniques and professionals involved in the diagnosis make it important to design a platform that covers the entire problem. DERMA was born to achieve this objective using all the features extracted from the preliminary analysis of the problem. The platform is a decision support system to help medical experts in their diagnosis. The general architecture permits the integration of several data sources that correspond to the different medical profiles and techniques involved in a melanoma diagnosis. Each data source is used to configure a single CBR subsystem that performs a single classification. The combination of all CBR subsystems through a medical protocol scheme results in a final collaborative diagnosis.

Every challenge we have addressed has improved the results of previous steps. We started our work with a single classification, using a single CBR system with just one kind of data. This first step allows us to determine the base accuracy. Later, we propose a collaborative diagnosis following the medical protocol, where different subsystems make a unified prognostic with different data sources. This module performs the same process used by medical experts combining different criteria. The basic collaboration was followed with the application of enhancements on the knowledge base organization and the collaborative process. This work tunes the system in order to improve the sensitivity and specificity rates. Finally, we extended the work to the use of multilabel data due to the domain characteristics. This last step offers good results and keeps the door open to richer datasets. We could summarize that the results obtained by DERMA during the test process were the ones expected by medical experts. Nowadays the most outstanding problem remains to be the lack of enough melanoma data, but we are developing a data managing application that will solve this problem. Once we obtain the complete data that fits all the melanoma features, we will recheck our results in order to use DERMA system to help in the day-to-day assistance which is our main future goal.

Moreover, analog reasoning and hybrid systems, such as collaborative and multilabel CBR, are hot research topics particularly in the area of health and medicine. It is crucial to have a framework as flexible as the one offered by this family of techniques. They are highly reliable within the community of data mining. Some of these processes are being explored in new fields such as social networks and marketing areas. Techniques such as analog reasoning exploit the high capacity of the computer to find patterns that lead to new and useful information for the user. Our proposals could be moved to these new areas in order to take care of new problems. This is the second part of our further work.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

The authors thank the Agencia de Gestio d'Ajuts Universitaris i de Recerca for funding the Grup de Recerca en Sistemes Intel·ligents as recognized group (2009-SGR-183) and for Grant 2010FI_B2 0084.

References

[1] J. E. McWhirter and L. Hoffman-Goetz, "Visual images for patient skin self-examination and melanoma detection: a systematic review of published studies," *Journal of the American Academy of Dermatology*, vol. 69, no. 1, pp. 47–55, 2013.

[2] A. Aamodt and E. Plaza, "Case-based reasoning: foundational issues, methodological variations, and system approaches," *AI Communications*, vol. 7, no. 1, pp. 39–59, 1994.

[3] S. Segura, S. Puig, C. Carrera, J. Palou, and J. Malvehy, "Dendritic cells in pigmented basal cell carcinoma: a relevant finding by reflectance-mode confocal microscopy," *Archives of Dermatology*, vol. 143, no. 7, pp. 883–886, 2007.

[4] A. Scope, C. Benvenuto-Andrade, A.-L. C. Agero et al., "*In vivo* reflectance confocal microscopy imaging of melanocytic skin lesions: consensus terminology glossary and illustrative images," *Journal of the American Academy of Dermatology*, vol. 57, no. 4, pp. 644–658, 2007.

[5] J. Avila, E. Gibaja, and S. Ventura, "Multi-label classification with gene expression programming," in *Hybrid Artificial Intelligence Systems*, vol. 5572 of *Lecture Notes in Computer Science*, pp. 629–637, Springer, Berlin, Germany, 2009.

[6] P. E. Xing, Y. A. Ng, M. I. Jordan, and S. Russell, "Distance metric learning, with application to clustering with side-information," in *Advances in Neural Information Processing Systems 15*, pp. 505–512, The MIT Press, London, UK, 2002.

[7] A. Jain, A. Jain, and S. Jain, *Artificial Intelligence Techniques in Breast Cancer Diagnosis and Prognosis*, Series in Machine Perception and Artificial Intelligence, World Scientific, Hackensack, NJ, USA, 2000.

[8] N. Singh, A. G. Mohapatra, and G. Kanungo, "Breast cancer mass detection in mammograms using K-means and fuzzy C-means clustering," *International Journal of Computer Applications*, vol. 22, no. 2, pp. 15–21, 2011.

[9] C.-R. Nicandro, M.-M. Efrén, A.-A. María Yaneli et al., "Evaluation of the diagnostic power of thermography in breast cancer using bayesian network classifiers," *Computational and Mathematical Methods in Medicine*, vol. 2013, Article ID 264246, 10 pages, 2013.

[10] A. Fornells, J. M. Martorell, E. Golobardes, J. M. Garrell, and X. VilasÍs, "Patterns out of cases using kohonen maps in breast cancer diagnosis," *International Journal of Neural Systems*, vol. 18, no. 1, pp. 33–43, 2008.

[11] E. Armengol and S. Puig, "Combining two lazy learning methods for classification and knowledge discovery," in *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval (INSTICC '11)*, pp. 200–207, Paris, France, 2011.

[12] A. Fornells, E. Golobardes, E. Bernadó, and J. Martí-Bonmatí, "Decision support system for breast cancer diagnosis by a meta-learning approach based on grammar evolution," in *Proceedings of the 8th International Conference on Enterprise Information Systems (ICEIS '06)*, pp. 222–229, May 2006.

[13] A. Fornells, E. Armengol, E. Golobardes, S. Puig, and J. Malvehy, "Experiences using clustering and generalizations for knowledge discovery in melanomas domain," in *Advances in Data Mining. Medical Applications, E-Commerce, Marketing, and Theoretical Aspects*, vol. 5077 of *Lecture Notes in Computer Science*, pp. 57–71, Springer, Berlin, Germany, 2008.

[14] E. Armengol, "Classification of melanomas in situ using knowledge discovery with explained case-based reasoning," *Artificial Intelligence in Medicine*, vol. 51, no. 2, pp. 93–105, 2011.

[15] V. Ahlgrimm-Siess, M. Laimer, E. Arzberger, and R. Hofmann-Wellenhof, "New diagnostics for melanoma detection: from artificial intelligence to RNA microarrays," *Future Oncology*, vol. 8, no. 7, pp. 819–827, 2012.

[16] S. M. Rajpara, A. P. Botello, J. Townend, and A. D. Ormerod, "Systematic review of dermoscopy and digital dermoscopy/artificial intelligence for the diagnosis of melanoma," *British Journal of Dermatology*, vol. 161, no. 3, pp. 591–604, 2009.

[17] Y. Kalfoglou and M. Schorlemmer, "Ontology mapping: the state of the art," *The Knowledge Engineering Review*, vol. 18, no. 1, pp. 1–31, 2003.

[18] T. G. Dietterich, "Ensemble learning," in *The Handbook of Brain Theory and Neural Networks*, pp. 405–408, The MIT Press, Cambridge, Mass, USA, 2002.

[19] R. Nicolas, E. Golobardes, A. Fornells, S. Puig, C. Carrera, and J. Malvehy, "Identification of relevant knowledge for characterizing the melanoma domain," in *2nd International Workshop on Practical Applications of Computational Biology and Bioinformatics (IWPACBB '08)*, vol. 49 of *Advances in Soft Computing*, pp. 55–59, Springer, Berlin, Germany, 2009.

[20] J. Hartigan and M. Wong, "A k-means clustering algorithm," in *Applied Statistics*, vol. 28, pp. 100–108, 1979.

[21] T. Kohonen, *Self-Organizing Maps*, Springer Series in Information Sciences, Springer, New York, NY, USA, 3rd edition, 2000.

[22] D. Vernet, R. Nicolas, E. Golobardes et al., *Pattern Discovery in Melanoma Domain Using Partitional Clustering*, vol. 184 of *Frontiers in Artificial Intelligence and Applications*, IOS Press, Amsterdam, The Netherlands, 2008.

[23] A. Fornells, E. Golobardes, J. M. Martorell, J. M. Garrell, E. Bernado, and N. Macia, "A methodology for analyzing the case retrieval from a clustered case memory," in *Case-Based Reasoning Research and Development*, vol. 4626 of *Lecture Notes in Computer Science*, pp. 122–136, Springer, Berlin, Germany, 2007.

[24] S. Wess, K. D. Althoff, and G. Derwand, "Using *k*-d trees to improve the retrieval step in case-based reasoning," in *Topics in Case-Based Reasoning*, vol. 837 of *Lecture Notes in Artificial Intelligence*, pp. 167–181, Springer, Berlin, Germany, 1994.

[25] M. Lenz, H. D. Burkhard, and S. Brückner, "Applying case retrieval nets to diagnostic tasks in technical domains," in *Advances in Case-Based Reasoning*, vol. 1168 of *Lecture Notes in Artificial Intelligence*, pp. 219–233, Springer, Berlin, Germany, 1996.

[26] Q. Yang and J. Wu, "Enhancing the effectiveness of interactive case-based reasoning with clustering and decision forests," *Applied Intelligence*, vol. 14, no. 1, pp. 49–64, 2001.

[27] E. L. Rissland and J. J. Daniels, "The synergistic application of CBR to IR," *Artificial Intelligence Review*, vol. 10, no. 5-6, pp. 441–475, 1996.

[28] A. Fornells, E. Golobardes, D. Vernet, and G. Corral, "Unsupervised case memory organization: analysing computational time and soft computing capabilities," in *Advances in Case-Based Reasoning*, vol. 4106 of *Lecture Notes in Computer Science*, pp. 241–255, Springer, Berlin, Germany, 2006.

[29] A. Fornells, J. Camps, E. Golobardes, and J. M. Garrell, "Comparison of strategies based on evolutionary computation for the design of similarity functions," in *Artificial Intelligence Research and Development*, vol. 131, pp. 231–238, IOS Press, Amsterdam, The Netherlands, 2005.

[30] Y. Liu, Y. Yang, and J. Carbonell, "Boosting to correct inductive bias in text classification," in *Proceedings of the 11st International Conference on Information and Knowledge Management (CIKM '02)*, pp. 348–355, ACM Press, New York, NY, USA, November 2002.

[31] R. Nicolas, D. Vernet, E. Golobardes, A. Fornells, F. De la Torre, and S. Puig, "Distance metric learning in a collaborative melanoma diagnosis system with case-based reasoning," in *Proceedings of the 14th United Kingdom Workshop on Case-Based Reasoning at the 29th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence*, pp. 58–66, CMS Press, University of Greenwich, London, UK, 2009.

[32] E. Bauer and R. Kohavi, "An empirical comparison of voting classification algorithms: bagging, boosting, and variants," *Machine Learning*, vol. 36, no. 1-2, pp. 105–139, 1999.

[33] K. T. Leung and D. S. Parker, "Empirical comparisons of various voting methods in bagging," in *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '03)*, pp. 595–600, ACM Press, New York, NY, USA, 2003.

[34] R. Avogadri and G. Valentini, "Fuzzy ensemble clustering for DNA microarray data analysis," in *Applications of Fuzzy Sets Theory*, vol. 4578 of *Lecture Notes in Computer Science*, pp. 537–543, Springer, Berlin, Germany, 2007.

[35] R. E. Abdel-Aal, "Abductive network committees for improved classification of medical data," *Methods of Information in Medicine*, vol. 43, no. 2, pp. 192–201, 2004.

[36] S. Ontañón, *Ensemble Case-Based Learning for Multi-Agent Systems*, VDM, Saarbrücken, Germany, 2008.

[37] P. Melville and R. J. Mooney, "Diverse ensembles for active learning," in *Proceedings of the 21st International Conference on Machine Learning (ICML '04)*, pp. 584–591, Banff, Canada, July 2004.

[38] R. Nicolas, E. Golobardes, A. Fornells et al., *Using Ensemble-based Reasoning to Help Experts in Melanoma Diagnosis*, vol. 184 of *Frontiers in Artificial Intelligence and Applications*, IOS Press, Amsterdam, The Netherlands, 2008.

[39] R. Nicolas, D. Vernet, E. Golobardes, A. Fornells, S. Puig, and J. Malvehy, "Improving the combination of CBR systems with preprocessing rules in melanoma domain," in *Proceedings of the 8th International Conference on Case-Based Reasoning*, pp. 225–234, Seattle, Wash, USA, 2009.

[40] T. K. Ho, M. Basu, and M. Law, "Measures of complexity in classification problems," in *Data Complexity in Pattern Recognition*, Advanced Information and Knowledge Processing, pp. 1–23, Springer, London, UK, 2006.

[41] G. Tsoumakas and I. Vlahavas, "Random *K*-labelsets: an ensemble method for multilabel classification," in *Machine Learning: ECML 2007*, vol. 4701 of *Lecture Notes in Computer Science*, pp. 406–417, 2007.

[42] M.-L. Zhang and Z.-H. Zhou, "ML-KNN: a lazy learning approach to multi-label learning," *Pattern Recognition*, vol. 40, no. 7, pp. 2038–2048, 2007.

[43] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, The Morgan Kaufmann Series in Data Management Systems, Morgan Kaufmann, San Francisco, Calif, USA, 2006.

[44] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *SIGKDD Explorations*, vol. 11, no. 1, pp. 10–18, 2009.

[45] J. Read and A. Pruned, "Problem transformation method for multi-label classification," in *Proceedings of the 6th New Zealand Computer Science Research Student Conference (NZCSRSC' 08)*, pp. 143–150, Christchurch, New Zealand, 2008.

[46] M.-L. Zhang and Z.-H. Zhou, "Multilabel neural networks with applications to functional genomics and text categorization," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 10, pp. 1338–1351, 2006.

[47] R. E. Schapire and Y. Singer, "Boostexter: a boosting-based system for text categorization," *Machine Learning*, vol. 39, no. 2, pp. 135–168, 2000.

[48] J. Ross Quinlan, *C4.5: Programs for Machine Learning*, Morgan Kaufmann Series in Machine Learning, Morgan Kaufmann, San Francisco, Calif, USA, 1993.

[49] A. Clare and R. D. King, "Knowledge discovery in multi-label phenotype data," in *Principles of Data Mining and Knowledge Discovery*, vol. 2168 of *Lecture Notes in Computer Science*, pp. 42–53, Springer, Berlin, Germany, 2001.

[50] A. Elisseeff, J. Weston, and A. Kernel, "Method for multi-labelled classification," in *Advances in Neural Information Processing Systems*, vol. 14, pp. 681–687, The MIT Press, Cambridge, Mass, USA, 2001.

[51] R. Nicolas, A. Sancho-Asensio, E. Golobardes, A. Fornells, and A. O. Puig, "Multi-label classification based on analog reasoning," *Expert Systems with Applications*, vol. 40, no. 15, pp. 5924–5931, 2013.

[52] A. Frank and A. Asuncion, *UCI Machine Learning Repository*, University of California, Berkeley, Calif, USA, 2010.

Multi-label classification based on analog reasoning. *Expert Systems With Applications Journal*, 2013

Expert Systems with Applications 40 (2013) 5924–5931

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Multi-label classification based on analog reasoning

Ruben Nicolas *, Andreu Sancho-Asensio, Elisabet Golobardes, Albert Fornells, Albert Orriols-Puig

Grup de Recerca en Sistemes Intel.ligents, La Salle – Universitat Ramon Llull, Quatre Camins 2, 08022 Barcelona, Spain

ARTICLE INFO

Keywords:
Multi-label
Classification
Case-Based Reasoning

ABSTRACT

Some of the real-world problems are represented with just one label but many of today's issues are currently being defined with multiple labels. This second group is important because multi-label classes provide a more global picture of the problem. From the study of the characteristics of the most influential systems in this area, MlKnn and RAkEL, we can observe that the main drawback of these specific systems is the time required. Therefore, the aim of the current paper is to develop a more efficient system in terms of computation without incurring accuracy loss. To meet this objective we propose MlCBR, a system for multi-label classification based on Case-Based Reasoning. The results obtained highlight the strong performance of our algorithm in comparison with previous benchmark methods in terms of accuracy rates and computational time reduction.

1. Introduction

Recent progress in machine learning and data mining has led to the application of their techniques in more complex multi-label problems, such as forecasting, where we have data with different features obtained from several stations that could be used in order to predict just one class, e.g., rain probability, but other relevant classes could be analyzed together to provide a more global picture of the forecast. These labels could be temperature, humidity, wind, and so on. This difficulty is also emphasized in other categorization problems such as medicine, emotions, texts, biology, or face verification, among others because they are complex issues that could be analyzed from more than one point of view (Tsoumakas, Katakis, & Vlahavas, 2008). There are two ways to tackle this problem (Tsoumakas & Katakis, 2007): (1) to transform the dataset to single-label and use classical classification algorithms or (2) to modify these classical algorithms to accept multi-label data. In our case we have worked on the second given that the first family is somehow a step backwards towards the single-label classification because these systems lose the possibility of analyzing the problem from different points of view. Within this second group there are several contributions among which MlKnn (Zhang & Zhou, 2005, 2007) and RAkEL (Tsoumakas et al., 2007) are the most noteworthy. These two concrete proposals face the problem effectively in terms of accuracy but they are not efficient time-wise. This paper tackles the difficulty of reducing the computational cost of classification from multi-label data without losing the precision achieved with previous methods. This is really important from the standpoint that nowadays problems can be represented with datasets that are not only rich in labels but also in the number of cases. The increment of instances correlates a direct increase in computational time.

To achieve a reduction in time costs without penalizing the accuracy we propose a Case-Based Reasoning (CBR) (Aamodt & Plaza, 1994) system for multi-label classification based on MlKnn fundamentals. The choice of CBR as the core of our algorithm is based on its main skills: (1) good adaptation to multi-label characteristics; (2) low complexity being a competent method; (3) explicative capability of CBR that is extremely important in problems such as medical prognosis; (4) existences of an active CBR community that is interested in the adaptation to multi-label problems (Brinker & Hüllermeier, 2007); (5) non-existence of an approach to this goal using CBR despite the interest of the researchers in this area. In addition if we consider CBR as an improvement on Knn systems, we should also consider it as an effective approach to enhancing some of the characteristics of MlKnn. Our adaptation of CBR algorithm to multi-label problems has been focused on the retrieval and reuse stages. Results of our proposal are compared with other two competitive multi-label learning systems, MlKnn and RAkEL, using seven synthetic dataset and three other real-world datasets used as benchmark by multi-label classification community (Ávila, Gibaja, & Ventura, 2009). The algorithms are compared with Friedman, Holm and Shaffer statistical tests.

The remainder of this paper is organized as follows: Section 2 summarizes the background information and the related work; Section 3 presents the contribution for multi-label classification; Section 4 describes the experimentation and discusses the results; and finally, Section 5 ends with the conclusions and further work.

* Corresponding author. Tel.: +34 932902451.
E-mail address: rnicolas@salle.url.edu (R. Nicolas).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.05.004

2. Related work

This paper tackles the difficulty of reducing the computational cost of classifying using multi-label data without losing accuracy. Currently the work in multi-label problem solving is divided into two different families. On the one hand, Problem Transformation Methods (PTM) transform the learning task into one or more single-label classification tasks. The main problem of this family is that with the unification of different labels into a single one we may lose information that could be critical in cases such as medical prognosis. In contrast, the positive aspect is the possibility of using existing algorithms without having to modify them. On the other hand, Algorithm Adaptation Methods (AAM) deal with the problem of modifying classical algorithms to work in a multi-label mode. Despite the fact that this second family of methods focuses on not losing information, the adaptation is not trivial and could increase the calculations and, consequently, the time consume.

The most influential works from the AAM family include (Schapire, 2000) which presents the BoosTexter a system that uses boosting algorithms for text-categorization. This platform, designed for automatic call-type identification makes classes for further classification; (Clare & King, 2001) that deal with multi-label biological data and adapt the entropy analysis in order to use classical C4.5 (Quinlan, 1993) algorithm to create a decision tree; and (Elisseeff & Weston, 2001) that focuses its attention on an approach based on a ranking method combined with a predictor of the size of the sets which tries to overcome the difficulties found by previous works when adapting multi-label problems to two classes ones.

The most competent works in PTM for multi-label classification are MlKnn and RAkEL. These two works are recognized by the community as reference algorithms. The first one, MlKnn, is a theoretical approach to multi-label classification that adapts the combination of the k recovered cases of classical k nearest-neighbor algorithm (Knn) (Han & Kamber, 2006) to multiple label problems. RAkEL is an ensemble platform that allows the classification of multi-label datasets by dealing with each label separately and combining the single-label results. It can be used with several algorithms as a single-label classifier system (as it is implemented using WEKA (Hall et al., 2009) libraries all its classifiers can be used) but the one used as a common benchmark is C4.5. Both algorithms are publicly available with a standard configuration under the name of MULAN. Although MlKnn and RAkEL are the most competent and commonly used platforms there are other interesting works in this field used by the community such as (Read, 2008) where the authors present a pruned transformation that combines key-points of several previous approaches and (Zhang & Zhou, 2006) that uses neural networks for multi-label classification.

In reference to the characteristics of previous works in multi-label classifications and its shortcomings, we have developed our proposals based on AAM because, attending to the literature, this family reaches better results than PTM. These are described in the following section.

3. Multi-label Case-Based Reasoning Algorithm

Current multi-label classification methods in the AAM family provide competent accuracy results but show high complexity in terms of computation. These systems propose a complex algorithm with a high level of calculus that increases the computational time. Our proposal obtains a system which is as accurate as previous ones but which employs less calculus and is, therefore, less complex. This platform has been named Multi-label Case-Based Reasoning (MlCBR). The most similar work that addresses the use of algorithms with small number of calculations for multi-label classification is MlKnn. This work proposes the adaptation of Knn algorithm to multi-label classification. The changes suggested by Zhang and Zhou (2005) to transform the single-label algorithm into a multi-label approach are the addition of some mathematical calculations after recovering the k most similar cases of the case memory. In our case, unlike MlKnn we adapted to multi-label classification by employing CBR method, which is a technique that solves new cases by using others previously solved. In order to achieve this objective, four phases are applied: (1) first of all, the system retrieves the most similar cases from the case memory with the assistance of a similarity function; (2) secondly, it tries to reuse the solutions from the retrieved cases with the aim to solve the present case, (3) then it revises the solution, and (4) finally it retains the useful information of the solved case, if necessary. All the steps are centered on the case memory, which contains the experience of system in terms of cases. A case is an instance of a problem. We have chosen this algorithm because to a certain extent it is an improvement on Knn by the addition of the retaining, revising and reuse phases to the simple retrieval of the other option. Furthermore, the competence of this kind of algorithm is visible in problems related to medicine, semantic web or general purpose classification. The main advantages of CBR that make it perfect for a multi-label transformation are its accredited results of good performance and low complexity, its explicative capacity and the fact that it has an active community working on it which is interested in this specific kind of problems. In this paper we centered our effort on the retrieve and reuse stages of CBR because these are the features that will enable us to meet our objectives, namely the reduction of computational time and maintaining or improving the accuracy. The retrieve stage of Multi-label Case-Based Reasoning Algorithm (MlCBR) algorithm is based on MlKnn where the k most similar cases to the case study are recovered of the case memory. As reuse phase two approaches are proposed. Probabilistic Reuse (PR) is the first option where the final classification is made through a voting process which all the recovered cases are equally weighted. In contrast, Probabilistic Reuse based on Experience (PRE) adds the concept of experience to better weight the recovered cases. In the following subsections we detail the algorithms proposed for reuse stage on multi-label classification using CBR: PR and PRE.

3.1. First step: Probabilistic Reuse

Probabilistic Reuse algorithm present probabilistic variations in the classical reuse stage in order to adapt CBR to multi-label classification. Once the system recovers the k best cases, they are mixed in order to propose a solution. We consider a voting combination of cases similar to the one proposed by MlKnn but adapted to the CBR idea. MlKnn, in the same way as other single-label Knn algorithms, recovers the k best cases of the previously recovered ones and gives a classification result combining the k cases. This combination is done through counting the number of recovered instances that predict each label. The platform considers that a label will be set to one if more than a half of the k recovered cases have this label with a positive value. In the case of MlCBR with PR reuse, after we recover the k best cases in the retrieval stage we combine it in reuse. This reuse step considers the frequency for each label and sets it to positive if the percentage is more than 50%. The mathematical process followed by MlKnn and PR reuse to combine the k cases obtained is the same in terms of the final result. The

description of MlKnn provided by Zhang and Zhou is done through mathematical formulation and not algorithmically. During the implementation of MlCBR retrieve and PR reuse we make the most of the programming language potential reducing the calculations performed for each comparison and moreover decreasing the number of comparisons proposed by MlKnn. The improvements in time consume can be seen in Section 4. These differences made MlCBR a less complex algorithm in comparison with MlKnn in terms of calculations. In addition, it is a more modular system, which will allow easy changes in future works without modifying the core system. The process followed is described by Algorithm 1. As we can see, the system recovers the k best cases and explores them for each label updating the percentage of appearances. Finally the system decides if the label is positive or not through a threshold value (50% in this case).

Algorithm 1. MlCBR Classification Algorithm using PR.

```
Let c_new be the case study
Let c_mem be the case memory
Let k be the number of cases to consider as best ones
Let l be the number of labels
Let k_best be the k best cases from c_mem
Let l_sum be the proportion of appearance for each label
Let res be the predicted value for each label
k_best ← recoverKbest(k, c_mem, c_new)
foreach i ← 1 to l do
    foreach j ← 1 to k do
        l_sum[i] ← l_sum[i] + ((k_best[j][i]) / k)
    i ← i + 1
foreach i ← 1 to l do
    if l_sum[i] > 0.5 then
        res[i] ← 1
    else
        res[i] ← 0
return res
```

3.2. Second step: improving Probabilistic Reuse using experience

There are different works that add the experience concept to CBR classification. These works underscore the importance of weighting the retrieved cases in order to achieve better accuracy results. The major approach to this technique is the use of CBR with some changes inspired by reinforcement learning (Sutton & Barto, 1998; Salamo & Golobardes, 2004). The case of our experience based reuse, named PRE, adds a positive feedback to the recovered classification if it is correct and a negative one otherwise. This feedback will be used in further classifications in order to take a determined recovered case into higher or lower consideration. The combination algorithm of PR reuse is then modified in order to weigh up some of the k recovered cases during the ensemble voting protocol from previous classifications. This process has two parts: first of all, the case memory has an associated value for each case that considers the experience obtained by the case in previous classifications. This value is updated for each new case classification following Algorithm 2. When this case is recovered, and it contributes to the final classification, if the classification is correct a positive feedback is given. Otherwise, a negative reward is associated to the cases that cause the misclassification. The value associated to this experience is previously set. Secondly, during the process of classification of a new case, the algorithm multiplies the proportion of appearance of each label by experience (feedback) of the recovered case. This ponderation process is done in order to take into higher or lower consideration the recovered cases during the ensemble process. It is described by Algorithm 3.

Algorithm 2. MlCBR Experience Update Algorithm using PRE.

```
Let c_new be the case study
Let c_new_labels be the real labels of the case study
Let k_best be the k best cases from c_mem
Let k_best_labels be the real labels of the k best cases from c_mem
Let res be the predicted value for each label
Let k be the number of cases to consider as best ones
Let c_mem be the case memory
Let c_mem_exp be the experience value for each case of the case memory
Let exp be the fixed percentage of experience
if c_new_labels = res then
    foreach i ← 1 to k do
        if k_best_labels[i] = res then
            c_mem_exp[k_best] + exp
else
    foreach i ← 1 to k do
        if k_best_labels[i] = res then
            c_mem_exp[k_best] − exp
return c_mem
```

Algorithm 3. MlCBR Classification Algorithm using PRE.

```
Let c_new be the case study
Let c_mem be the case memory
Let c_mem_exp be the experience value for each case of the case memory
Let k be the number of cases to consider as best ones
Let l be the number of labels
Let k_best be the k best cases from c_mem
Let l_sum be the proportion of appearance for each label
Let res be the predicted value for each label
k_best ← recoverKbest(k, c_mem, c_new)
foreach i ← 1 to l do
    foreach j ← 1 to k do
        l_sum[i] ← l_sum[i] + (((k_best[j][i]) / k) · (c_mem_exp[k_best][k]))
    i ← i + 1
foreach i ← 1 to l do
    if l_sum[i] > 0.5 then
        res[i] ← 1
    else
        res[i] ← 0
return res
```

Having described the classification algorithm and its different options for reuse we can now describe the experimentation performed and its results. It is done in the following section.

4. Experiments, results and discussion

This section describes the experimentation performed in this work. Firstly we introduce the datasets we used and their characteristics. Secondly we highlight the steps followed during the experimentation, the algorithms involved, and the type of test and statistical comparison methods. Thirdly, we show the results obtained with our platform MlCBR (with PR and PRE reuse stages) and the other state of the art ones. Finally, we discuss the results and compare the platforms.

4.1. Test bed

We have used ten datasets (described in Table 1) for our experimentation where three are real-world problems used by the researchers that designed other competent multi-label classification algorithms and seven are problems which are synthetic whose datasets generation algorithm was proposed by ourselves. The choice of these datasets and the need to propose synthetic data is conditioned by the characteristics of the domain and the lack of available data. In contrast to single-label classification tasks,

Table 1
Summary of the properties of the dataset used. The columns include: identification (dataset), type of data (domain), number of instances (instances), number of attributes (attributes), and number of labels (labels).

Dataset	Domain	Instances	Attributes	Labels
Multilabel2D	Synthetic	3000	2	3
Multilabel3D	Synthetic	3000	3	3
Multilabel4D	Synthetic	3000	4	3
Multilabel5D	Synthetic	3000	5	3
Multilabel6D	Synthetic	3000	6	3
Multilabel7D	Synthetic	3000	7	3
Multilabel8D	Synthetic	3000	8	3
Scene	Image	2407	294	6
Emotions	Music	593	72	6
Yeast	Biology	2417	103	14

where there are large amounts of representative datasets as is the case of the UCI repository (Frank & Asuncion, 2010) to evaluate and compare different algorithmic approaches, this is not the case of the multi-label field. The majority of works (Ávila et al., 2009) use three datasets (scene, emotions and yeast) in order to compare different approaches. Thus, our reasons for designing a synthetic multi-label dataset generator are the following: (1) develop a common benchmark, independent of the machine learning method used, (2) be simple enough to let researchers study the results in great detail but be complete enough to become representative, (3) allow researchers to modify the complexity of the problem by changing the number of variables in order to analyze how this affects the methods used, and (4) be 100% replicable. With these ideas in mind we wanted a reliable environment to compare different algorithms and to provide a test bed for future works related to multi-label classification tasks. To simplify the description we introduced a two-dimensional problem generated by our synthetic dataset generator in as explained below.

The dataset consists of three hyper spheres (each one represents a separate central label) that are partially overlapped. These hyper spheres are defined by two continuous variables $\{x_1,x_2\}$ ranging in $[0,1]$ and following a Normal distribution $N(\frac{a_i}{10},\frac{1}{10})$, where a_i is a continuous variable that differs for each x_i and for each hyper sphere in the following way: $a_1 = 4$ for x_1 and $a_2 = 3$ for x_2 in the first hyper sphere, $a_1 = 6$ for x_1 and $a_2 = 3$ for x_2 in the second hyper sphere, and finally $a_1 = 5$ for x_1 and $a_2 = 6$ for x_2 in the third hyper sphere.

Each example has a set of three labels associated $\{\ell_1,\ell_2,\ell_3\}$, where each ℓ_i can take the values $\{0,1\}$ depending on the region the example is set in. Thus, if an example is inside the first hyper sphere in the region where there is no overlap with either of the other two, the labels will be set as $\ell_1 = 1$, $\ell_2 = 0$, $\ell_3 = 0$. If the example lies in the overlapped region among hyper spheres two and three, the labels of this one will be set as $\ell_1 = 0$, $\ell_2 = 1$, $\ell_3 = 1$. To obtain this we use the following formula:

$$\ell_i \leftarrow \begin{cases} 1 & \text{if } \left(x_1 - \frac{a_1}{10}\right)^2 + \left(x_2 - \frac{a_2}{10}\right)^2 \leqslant radius^2, \\ 0 & \text{otherwise}, \end{cases} \quad (1)$$

where *radius* is the actual radius of the hyper sphere, set to 0.25. Thus, formula 1 is used three times for each example: one for each hyper sphere to properly set the labels of each example. The total number of examples is, as default, 3000, that is, 1000 examples for each hyper sphere.

In order to check the response of the algorithms in high dimensional spaces, this synthetic dataset has extended to n-dimensions as follows: the number of input variables is extended to the desired dimension (i.e., if we want to use five dimensional hyper spheres, we will have the input variables $\{x_1,x_2,x_3,x_4,x_5\}$). The different a_i values keep the same as in the two dimensional case, repeating through the different variables, that is, $a_1 = 4$ for x_1, $a_2 = 3$ for x_2, then $a_3 = 4$ for x_3, $a_4 = 3$ for x_4, and so on for the first hyper sphere. The same with the second and third hyper spheres. The labels remain the same $\{\ell_1, \ell_2, \ell_3\}$, and to know the values of those we use the following formula:

$$\ell_i \leftarrow \begin{cases} 1 & \text{if } \sum_{i=1}^{n}\left(x_i - \frac{a_i}{10}\right)^2 \leqslant radius^2, \\ 0 & \text{otherwise}. \end{cases} \quad (2)$$

4.2. Experimental methodology

Initially, we set up the different parameters of our proposals. We test the performance of our platforms with different configurations and compare the results using Friedman, Holm and Shaffer statistical tests as recommended in Demšar (2006) in order to analyze the influence of k and *experience* values. The k values used for MlCBR (both for PR and PRE reuse stages) are $3, 5, 7, 9, 11, 13$, and 15. In the case of PRE we set the *experience* to $3\%, 5\%, 8\%$, and 15%. This values has been selected as a representative window from a larger set that we have previously tested. Once the set up is completed, we test the same dataset pool with MlKnn and RAkEL systems. All the phases of the experimentation process obtain the accuracy values with an average of 10 independent executions of ten-fold cross-validation process with different randomness seed. Furthermore, we save the computational time used for each dataset and algorithm in order to compare the time used for each proposal. The final results of MlCBR, MlKnn and RAkEL are compared using the same statistical test as in the case of setting up k and *experience* parameters.

4.3. MlCBR results

The first part of the experiment focuses on the effects of varying the configuration parameters, as well as reviewing the internal quality of the method. To obtain this goals we use the accuracy results for each setting in comparison with the other ones. First of all, we will see the results obtained using MlCBR algorithm with the base reuse stage, PR. These are summarized in Table 2 where we can see a first column with the dataset we are studying and the following ones with the results in percentage of the average accuracy and standard deviation of the system for each k value following the experimental methodology cited in the previous section. A preliminary analysis of the results shows that the accuracy results are lower with small k values than with medium ones. When k increases too much accuracy decreases. Taking a second step, and analyzing the results from an statistical point of view we applied the statistical tests previously cited. The statistical analysis of the results did not permit us to reject the null hypothesis that all configurations performed the same, on average, at $\alpha = 0.05$. Despite this, the ranking of algorithm configurations shows that $k = 13$ and $k = 9$ obtained the highest average accuracy.

The MlCBR algorithm with PRE reuse average accuracy results are shown in Tables 3–6. Each of these tables shows the datasets used for the experimentation and the different k values used. The difference between them is the experience value that is set to 0.03, 0.05, 0.08 and 0.15 respectively. The first analysis which refers just to the raw results, shows that MlCBR using PRE reuse follows the same curve as using PR where the medium values of k are the best ones. Besides, we can observe that small values of experience are positive in terms of accuracy in contrast with higher values. Studying table by table we can stress some results. In Table 3 we can see that the better results are obtained with k value set to 11, this is confirmed by ranking results. In comparison with PR (see Table 2) the results obtained by PRE with *experience* 3% are

Table 2
Comparison table of the average test performance and standard deviation of the ten times ten-fold cross-validation obtained by MlCBR using PR reuse.

Dataset	k03	k05	k07	k09	k11	k13	k15
Multilabel2D	96.87 ± 1.68	97.47 ± 0.65	97.13 ± 1.19	96. 93 ± 1.12	96.13 ± 1.02	95.80 ± 1.81	96.33 ± 1.69
Multilabel3D	93.33 ± 1.29	93.70 ± 1.68	94.00 ± 1.14	94. 30 ± 0.79	94.37 ± 0.94	95.17 ± 1.02	94.87 ± 0.92
Multilabel4D	79.07 ± 3.12	79.87 ± 2.93	80.47 ± 2.09	80. 13 ± 2.86	81.07 ± 2.91	80.37 ± 3.23	79.80 ± 3.31
Multilabel5D	70.73 ± 3.10	71.93 ± 2.74	72.13 ± 3.60	72. 97 ± 2.72	73.17 ± 2.99	73.57 ± 4.10	73.37 ± 2.03
Multilabel6D	84.80 ± 1.24	84.57 ± 1.91	84.73 ± 3.19	84. 83 ± 2.30	84.57 ± 2.80	84.57 ± 1.05	84.40 ± 2.31
Multilabel7D	81.50 ± 1.84	82.57 ± 3.35	82.20 ± 1.75	82. 87 ± 1.87	82.37 ± 2.56	82.10 ± 2.82	81.47 ± 2.60
Multilabel8D	81.20 ± 3.26	81.87 ± 2.03	83.13 ± 1.69	83. 37 ± 2.02	84.23 ± 3.48	84.27 ± 3.77	83.80 ± 1.80
Scene	67.43 ± 2.91	66.10 ± 3.00	65.48 ± 3.57	64.73 ± 3. 71	64.16 ± 3.80	63.95 ± 4.28	63.82 ± 3.51
Emotions	77.56 ± 5.10	77.59 ± 4.62	78.73 ± 5.63	78. 24 ± 5.89	78.42 ± 5.68	78.74 ± 5.75	78.91 ± 5.00
Yeast	98.30 ± 0.65	98.51 ± 0.59	98.55 ± 0.59	98.55 ± 0. 59	98.59 ± 0.53	98.59 ± 0.53	98.59 ± 0.53

Table 3
Comparison table of the average test performance and standard deviation of the ten times ten-fold cross-validation obtained by MlCBR using PRE reuse stage with 0.03 experience.

Dataset	k03	k05	k07	k09	k11	k13	k15
Multilabel2D	96.87 ± 1.68	97.27 ± 0.91	97.20 ± 1.45	97. 00 ± 1.20	96.13 ± 1.02	95.80 ± 1.81	96.33 ± 1.69
Multilabel3D	93.33 ± 1.29	93.27 ± 1.21	93.80 ± 0.80	93. 97 ± 1.00	94.37 ± 0.93	95.17 ± 1.02	94.93 ± 0.85
Multilabel4D	79.27 ± 2.61	79.87 ± 2.93	80.77 ± 3.22	80. 33 ± 2.88	81.07 ± 2.91	80.40 ± 3.22	79.80 ± 3.31
Multilabel5D	70.73 ± 3.10	72.23 ± 3.62	72.63 ± 3.04	72. 97 ± 2.72	73.17 ± 2.99	73.57 ± 4.10	73.37 ± 2.03
Multilabel6D	84.77 ± 2.20	84.17 ± 2.48	84.53 ± 2.09	84. 83 ± 2.30	84.57 ± 2.80	84.57 ± 1.05	84.33 ± 2.31
Multilabel7D	81.43 ± 1.72	82.80 ± 1.60	83.00 ± 1.85	82. 87 ± 1.87	82.37 ± 2.56	82.07 ± 2.81	81.40 ± 2.62
Multilabel8D	81.10 ± 1.55	82.53 ± 1.32	82.80 ± 1.83	83. 37 ± 2.02	84.23 ± 3.48	84.27 ± 3.77	83.83 ± 1.82
Scene	67.43 ± 2.91	66.10 ± 3.00	65.52 ± 3.56	64. 82 ± 3.73	64.24 ± 3.72	63.95 ± 4.29	63.99 ± 3.78
Emotions	77.90 ± 5.17	77.76 ± 4.72	78.56 ± 5.62	78. 07 ± 5.92	78.08 ± 5.74	78.74 ± 5.75	79.07 ± 5.19
Yeast	98.30 ± 0.65	98.51 ± 0.59	98.55 ± 0.59	98. 55 ± 0.59	98.59 ± 0.53	98.59 ± 0.53	98.59 ± 0.53

Table 4
Comparison table of the average test performance and standard deviation of the ten times ten-fold cross-validation obtained by MlCBR using PRE reuse stage with 0.05 experience.

Dataset	k03	k05	k07	k09	k11	k13	k15
Multilabel2D	96.87 ± 1.68	97.27 ± 0.91	97.20 ± 1.45	97. 00 ± 1.20	96.07 ± 1.00	96.00 ± 1.78	96.53 ± 1.80
Multilabel3D	93.33 ± 1.29	93.27 ± 1.21	93.80 ± 0.80	93. 97 ± 1.00	94.40 ± 0.91	95.13 ± 1.12	94.83 ± 0.82
Multilabel4D	79.27 ± 2.61	79.87 ± 2.93	80.77 ± 3.22	80. 30 ± 2.86	81.00 ± 2.97	80.40 ± 3.19	79.93 ± 3.36
Multilabel5D	70.73 ± 3.10	72.23 ± 3.62	72.63 ± 3.04	72. 97 ± 2.72	73.13 ± 2.96	73.60 ± 4.02	73.50 ± 2.03
Multilabel6D	84.77 ± 2.20	84.17 ± 2.48	84.53 ± 2.09	84. 83 ± 2.30	84.57 ± 2.80	84.40 ± 0.94	84.13 ± 2.27
Multilabel7D	81.43 ± 1.72	82.80 ± 1.60	83.00 ± 1.85	82. 87 ± 1.87	82.30 ± 2.48	81.93 ± 2.79	81.13 ± 2.90
Multilabel8D	81.10 ± 1.55	82.53 ± 1.32	82.80 ± 1.83	83. 33 ± 1.97	84.27 ± 3.47	84.27 ± 3.69	83.80 ± 1.79
Scene	67.43 ± 2.91	66.22 ± 3.01	65.61 ± 3.50	64.69 ± 3. 80	64.32 ± 3.70	63.91 ± 4.35	64.15 ± 3.95
Emotions	77.90 ± 5.17	77.76 ± 4.72	78.40 ± 5.54	77. 74 ± 5.98	78.41 ± 5.45	79.41 ± 5.85	78.91 ± 5.75
Yeast	98.30 ± 0.65	98.51 ± 0.59	98.55 ± 0.59	98.55 ± 0. 59	98.59 ± 0.53	98.59 ± 0.53	98.59 ± 0.53

Table 5
Comparison table of the average test performance and standard deviation of the ten times ten-fold cross-validation obtained by MlCBR using PRE reuse stage with 0.08 experience.

Dataset	k03	k05	k07	k09	k11	k13	k15
Multilabel2D	96.87 ± 1.68	97.27 ± 0.91	97.20 ± 1.45	97. 00 ± 1.20	96.00 ± 1.07	95.93 ± 1.79	96.33 ± 1.69
Multilabel3D	93.33 ± 1.29	93.27 ± 1.21	93.83 ± 0.84	93. 97 ± 1.00	94.43 ± 0.80	95.13 ± 1.00	94.90 ± 0.84
Multilabel4D	79.27 ± 2.61	79.87 ± 2.93	80.77 ± 3.22	80. 27 ± 2.87	80.87 ± 3.08	80.33 ± 3.33	79.93 ± 3.36
Multilabel5D	70.73 ± 3.10	72.23 ± 3.62	72.63 ± 3.04	72. 93 ± 2.80	73.07 ± 2.95	73.67 ± 4.08	73.50 ± 2.15
Multilabel6D	84.77 ± 2.20	84.17 ± 2.48	84.53 ± 2.09	84. 83 ± 2.21	84.40 ± 2.75	84.37 ± 1.05	83.93 ± 2.29
Multilabel7D	81.43 ± 1.72	82.80 ± 1.60	82.93 ± 1.87	82. 87 ± 1.86	82.33 ± 2.52	81.83 ± 2.48	80.87 ± 3.13
Multilabel8D	81.10 ± 1.55	82.53 ± 1.32	82.80 ± 1.83	83. 30 ± 1.92	84.30 ± 3.44	83.97 ± 3.73	83.57 ± 1.64
Scene	67.47 ± 2.96	66.26 ± 3.01	65.65 ± 3.55	64.98 ± 3. 88	64.70 ± 3.90	64.82 ± 4.27	64.61 ± 3.58
Emotions	77.90 ± 5.17	77.60 ± 4.87	78.57 ± 5.70	78. 25 ± 5.91	78.41 ± 5.50	79.41 ± 6.09	79.25 ± 5.74
Yeast	98.30 ± 0.65	98.51 ± 0.59	98.55 ± 0.59	98.55 ± 0. 59	98.59 ± 0.53	98.59 ± 0.53	98.59 ± 0.53

quite similar. Making the same analysis with *experience* set to 0.05 we can see that *k* value 11 is the best ranked (see Table 4). As in the previous case in comparison to PR the difference is not particularly big. With the increase of the importance of *experience* value we can see from Table 5 that with a value of 0.08 we need to increase *k* to 13 (instead of 11 which provides better results on *experience* 0.03 and 0.05) in order to obtain similar results. This is the first case where the accuracy decreases due to the *experience* value. As we can see in following sections this decrease, like the previous increases, is not significant according to the Holm/Shaffer test at $\alpha = 0.05$. Table 6 shows that the results obtained with *experience* 0.15 follow the same trend as *experience* 0.08 where accuracy is penalized. After the raw analysis of each *experience* value of PRE reuse, if we apply the same statistical procedure used with MlCBR

Table 6
Comparison table of the average test performance and standard deviation of the ten times ten-fold cross-validation obtained by MlCBR using PRE reuse stage with 0.15 experience.

Dataset	k03	k05	k07	k09	k11	k13	k15
Multilabel2D	96.87 ± 1.66	97.33 ± 0.78	97.07 ± 1.27	96. 53 ± 0.98	96.07 ± 1.13	96.07 ± 1.72	95.87 ± 1.97
Multilabel3D	93.33 ± 1.29	93.37 ± 1.84	94.07 ± 1.21	94. 03 ± 1.22	94.43 ± 0.36	95.10 ± 1.14	94.67 ± 0.81
Multilabel4D	79.13 ± 2.92	79.60 ± 3.45	80.53 ± 1.80	80. 13 ± 3.02	80.70 ± 3.13	80.40 ± 3.67	80.00 ± 3.09
Multilabel5D	70.73 ± 3.10	71.43 ± 3.15	72.40 ± 3.45	73. 03 ± 3.11	72.93 ± 2.93	73.80 ± 4.03	73.77 ± 2.10
Multilabel6D	84.73 ± 1.29	84.07 ± 2.08	84.80 ± 2.76	84. 47 ± 2.45	84.13 ± 2.71	84.07 ± 1.27	83.73 ± 2.12
Multilabel7D	81.40 ± 1.73	82.30 ± 3.59	81.67 ± 1.57	82. 20 ± 1.66	82.27 ± 2.66	81.60 ± 2.76	80.50 ± 3.05
Multilabel8D	81.00 ± 3.18	81.73 ± 2.04	82.97 ± 2.06	82. 63 ± 1.90	83.80 ± 3.59	84.03 ± 3.69	83.40 ± 1.62
Scene	67.59 ± 2.96	66.18 ± 2.99	65.35 ± 3.37	64.98 ± 4. 02	64.73 ± 3.90	65.07 ± 4.26	64.94 ± 3.42
Emotions	77.74 ± 5.16	77.44 ± 5.00	78.75 ± 5.87	77. 56 ± 6.17	77.06 ± 5.52	78.06 ± 5.45	78.59 ± 5.21
Yeast	98.30 ± 0.65	98.51 ± 0.59	98.59 ± 0.53	98.59 ± 0. 53	98.59 ± 0.53	98.59 ± 0.53	98.59 ± 0.53

with PR we can see that the difference between the configurations of PRE is not significant but $k = 11$ with *experience* = 0.05 and $k = 11$ with *experience* = 0.03 are the best ranked. These previous statistical results, which were used just to set up the proposed system, are not shown for brevity and because they are not relevant for the global method comparison.

4.4. State of the art results, discussion and comparison

The second set of outcomes analyzes the results obtained by the reference methods in multi-label classification and compares them with those of MlCBR. This comparison is done in order to prove that MlCBR is as good as the most important algorithms in this area and is also faster. Because these targets we have two experimental units. On the one hand, we analyze the competence of the method by means of the comparison of the accuracy results from the different methods, the evaluation of the error in 10 times 10-fold cross-validation and the use of statistical tests to explain the differences in accuracy rates. On the other hand we analyze the time consume from each system to show that MlCBR is faster than other state of the art platforms. The results obtained from MlKnn and RAkEL multi-label classification systems can be observed in Table 7. As in previous algorithms, the Table describes the dataset used for experimentation but in this case we have the average accuracy results for each system not related to k. This difference is because the public platform we used to test the algorithms only offers a standard configuration of the system. The application of a statistical test intensifies the fact that both methods are statistically equivalent (for $\alpha = 0.05$ and also for $\alpha = 0.10$) and that MlKnn is better ranked than RAkEL. Analyzing the results shown on Table 7 one by one we can observe that both MlKnn and RAkEL perform better results classifying synthetic datasets rather than using real world ones. This is a generalized trend in all the algorithms tested. This tendency could be explained through the number of attributes and labels of the datasets (see Table 1). The type of data used affects differently each algorithm although follows a common pattern.

Table 7
Comparison table of the average test performance and standard deviation of the ten times ten-fold cross-validation obtained by MlKnn and RAkEL.

Dataset	MlKnn	RAkEL
Multilabel2D	96.85 ± 1.51	95.13 ± 1.52
Multilabel3D	94.33 ± 1.02	90.63 ± 1.79
Multilabel4D	81.28 ± 1.99	78.78 ± 2.15
Multilabel5D	74.06 ± 2.27	70.00 ± 2.42
Multilabel6D	85.10 ± 2.08	80.99 ± 2.15
Multilabel7D	84.59 ± 2.08	77.20 ± 2.20
Multilabel8D	86.15 ± 1.76	78.88 ± 2.39
Scene	8.59 ± 0.76	56.56 ± 2.85
Emotions	19.36 ± 2.21	25.99 ± 4.88
Yeast	19.26 ± 0.87	11.29 ± 1.96

Table 8
Friedman's average rank table and the position in this ranking of the best ranked algorithms.

Algorithm	Ranking	Position
MlKnn	2.80	1
MlCBR (PRE k11 exp05)	2.90	2
MlCBR (PR k13)	2.95	3
MlCBR (PRE k11 exp03)	3.25	4
MlCBR (PR k09)	3.30	5
RAkEL	5.80	6

Table 9
Holm/Shaffer table for $\alpha = 0.05$ of the best ranked algorithms.

i	Algorithms	z	p	Holm	Shaffer
15	**MlKnn vs. RAkEL**	**3.586**	**$3.362 \cdot 10^{-4}$**	**0.003**	**0.003**
14	**k11exp05 vs. RAkEL**	**3.466**	**$5.279 \cdot 10^{-4}$**	**0.004**	**0.005**
13	**k13 vs. RAkEL**	**3.406**	**$6.583 \cdot 10^{-4}$**	**0.004**	**0.005**
12	**k11exp03 vs. RAkEL**	**3.048**	**0.002**	**0.004**	**0.005**
11	**k09 vs. RAkEL**	**2.988**	**0.003**	**0.005**	**0.005**
10	k09 vs. MlKnn	0.598	0.550	0.005	0.005
9	k11exp03 vs. MlKnn	0.538	0.591	0.006	0.006
8	k09 vs. k11exp05	0.478	0.633	0.006	0.006
7	k13 vs. k09	0.418	0.676	0.007	0.007
6	k11exp03 vs. k11exp05	0.418	0.676	0.008	0.008
5	k13 vs. k11exp03	0.359	0.720	0.010	0.010
4	k13 vs. MlKnn	0.179	0.858	0.013	0.013
3	k11exp05 vs. MlKnn	0.120	0.905	0.017	0.017
2	k09 vs. k11exp03	0.060	0.952	0.025	0.025
1	k13 vs. k11exp05	0.060	0.952	0.050	0.050

At this point we have all the performance results of the different platforms separately. From these outcomes we can describe the global comparison of MlCBR with PR and PRE reuse, MlKnn, and RAkEL algorithms in terms of accuracy, computational time and statistical significance of the comparison. These systems are ranked as shown in Table 8 where we can see per columns the name of the algorithm tested, its ranking value and the position obtained. Even the ranking shows the better performance of MlKnn ahead of the best configurations of MlCBR with PR and PRE reuse stage while the statistical significance results draw attention to the difference is not statistically significant for $\alpha = 0.05$ as shown on Table 9. The difference is neither significant for $\alpha = 0.10$ whose results are not shown for brevity. This is not the case of RAkEL referring to the statistical results (see the results marked in bold in Table 9) have differences with MlCBR (with both reuse stages) and MlKnn that are statistically significant. These results related to the ranking (see Table 8) remark that RAkEL accuracy results are significantly worse than the other ones. From these results we can conclude that in terms of accuracy, our system (with PR

Table 10
Pairwise comparisons of the learning methods by means of a Holm's procedure. The below diagonal of the table shows the symbol ⊖ if the method in the row significantly degrades the method in the column at a significance level of 0.05 and +/− if there is no significant difference and performs better/worst.

	PR k09	PR k13	PRE k11 exp03	PRE k11 exp05	MlKnn	RAkEL
PR k09						
PR k13	+					
PRE k11 exp03	+	−				
PRE k11 exp05	+	+	+			
MlKnn	+	+	+	+		
RAkEL	⊖	⊖	⊖	⊖	⊖	

Table 11
Comparison table of the computational time (in seconds) used by RAkEL, MlKnn, and MlCBR algorithms during the classification process.

Dataset	RAkEL	MlKnn	MlCBR with PRE reuse	MlCBR with PR reuse
Synthetic data set	345600	274819	228246	225464
Real world data set	311040	157966	157784	155387
Total time	656640	432785	386030	380851
Percentage of improvement	–	34%	41%	42%

and PRE reuse) is equivalent to other competent platforms in this field such as MlKnn, and better than RAkEL, the benchmark system for multi-label classification. To summarize these results we can analyze Table 10 which makes a pairwise comparison of the learning methods by means of a Holm's procedure. Results should be analyzed by following the below diagonal of the table that shows the symbol ⊖ if the method in the row significantly degrades the method in the column at a significance level of 0.05; the + symbol if it performs better but with no significant difference; or the symbol − if, even with no significant difference, it performs worst. This pairwise comparison graphically shows that the performance of RAkEL is significantly worse in all cases while the results of the other methods tested are statistically equivalent. Furthermore, our system performs less calculus and it is less complex than the previous ones. For these reasons the system obtains the results in a shorter time and achieves higher modularity. The computational time used for experimentation is shown in Table 11. In addition, these results have a final row with the percentage of improvement for each system in comparison to the worst one in terms of computational time. These results have been obtained using the same computers and resources for each platform. If we contemplate these results we can underline that the time reduction between MlCBR with PR reuse and MlKnn is on average a 12% and 42% in comparison to RAkEL. This is interesting taking into consideration that today's problems deal with big datasets that require high computational time. The difference between the use of PRE and PR reuse stages is less than 1.5% and, as a consequence, MlCBR with PRE reuse is also better than RAkEL and MlKnn in terms of computation.

5. Conclusions and further work

In this paper we have discussed the need to obtain less complex classification systems for multi-label learning. This process should increase (or at least not decrease) the accuracy of previous systems, despite the reduction in complexity and in calculus. We can conclude that our proposal satisfies this criterion because, as experiments show, we obtain a level of accuracy equivalent to that obtained by a competent system (MlKnn) and statistically better results than the benchmark (RAkEL). In both comparisons the computational time of our proposals is lower than the one performed by previous platforms.

For further investigation, we will focus our efforts in the retain stage of CBR in order to improve the classification results. This improvement will be done using the feedback information obtained during PRE stage. Likewise we would like to establish patterns that allow for a better configuration of the system. In this line we would like to study the relation between the dataset and the obtained results. With this information we could choose a concrete configuration of our algorithms (or a concrete algorithm) according to the dataset we are dealing with. To face with this problem a multi-label data complexity measure should be proposed. Finally, new configurations of a multi-label synthetic dataset generator should be presented in order to propose a public multi-label dataset repository for future research in this area.

Acknowledgments

We would like to thank the Agencia de Gestio d'Ajuts Universitaris i de Recerca for funding the Grup de Recerca en Sistemes Intel.ligents as recognized group (2009-SGR-183) and for the FI Grant (2011FI_B 01028). Thanks must also go to Enginyeria La Salle and Universitat Ramon Llull for supporting our research group.

References

Aamodt, A., & Plaza, E. (1994). Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications, 7*, 39–59. ISSN: 0921-712.

Ávila, J., Gibaja, E., & Ventura, S. (2009). Multi-label classification with gene expression programming. In *Hybrid artificial intelligence systems. Lecture notes in computer science* (Vol. 5572, pp. 629–637). Berlin/Heidelberg: Springer. chap. 76.

Brinker, K., & Hüllermeier, E. (2007). Case-based multilabel ranking. In *Proceedings of the 20th international joint conference on artificial intelligence* (pp. 702–707). Morgan Kaufmann Publishers Inc..

Clare, A., & King, R. D. (2001). Knowledge discovery in multi-label phenotype data. In *Lecture notes in computer science* (pp. 42–53). Springer.

Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research, 7*, 1–30. ISSN: 1532-443.

Elisseeff, A., & Weston, J. (2001). A Kernel method for multi-labelled classification. *Advances in neural information processing systems* (Vol. 14, pp. 681–687). MIT Press.

Frank, A., & Asuncion, A. (2010). UCI machine learning repository, University of California, Irvine. <http://archive.ics.uci.edu/ml>, last checked: March 2011.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. In *SIGKDD explorations* (Vol. 11).

Han, J., & Kamber, M. (2006). Data mining: Concepts and techniques. In M. K. Publishers (Ed.), *The Morgan Kaufmann series in data management systems*, ISBN: 1-55860-901-6.

Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. Morgan Kaufman.

Read, J. (2008). A pruned problem transformation method for multi-label classification. In *Proceedings 2008 New Zealand computer science research student conference* (pp. 143–150).

Salamo, M., & Golobardes, E. (2004). Dynamic experience update for a case-based reasoning. In IOS Press (Ed.), *Frontiers in artificial intelligence and applications* (Vol. 109, pp. 158–164).

Schapire, R. E. (2000). Boostexter: A boosting-based system for text categorization. *Machine Learning, 39*, 135–168.

Sutton, R.-R., & Barto, A.-G. (1998). *Reinforcement learning: An introduction*. MIT Press.

Tsoumakas, G., & Vlahavas, I. (2007). Random k-labelsets an ensemble method for multilabel classification. In *Proceedings of the 18th European conference on machine learning* (pp. 406–417).

Tsoumakas, G., Katakis, I., & Vlahavas, I. (2008). Effective and efficient multilabel classification in domains with large number of labels. In *Proceedings of the*

European conference on machine learning and principles and practice of knowledge discovery in databases 2008 workshop on mining multidimensional data.

Tsoumakas, G., & Katakis, I. (2007). Multi-label classification: An overview. *International Journal of Data Warehousing and Mining, 3*, 1–13.

Zhang, M.-L., & Zhou, Z.-H. (2005). A k-nearest neighbor based algorithm for multi-label classification. In *Granular computing*. 0-7803-9017-2 (pp. 718–721). IEEE.

Zhang, M.-L., & Zhou, Z.-H. (2006). Multi-label neural networks with applications to functional genomics and text categorization. *IEEE Transactions on Knowledge and Data Engineering, 18*(10), 1338–1351.

Zhang, M.-L., & Zhou, Z.-H. (2007). Ml-Knn: A lazy learning approach to multi-label learning. *Pattern Recognition, 40*, 2038–2048.

Improving the combination of CBR systems with preprocessing rules in melanoma domain. *8th International Conference on Case-Based Reasoning, Workshop on Case-Based Reasoning in the Health Sciences*, 2009

Improving the Combination of CBR Systems with Preprocessing Rules in Melanoma Domain

Ruben Nicolas[1], David Vernet[1], Elisabet Golobardes[1], Albert Fornells[1], Susana Puig[2], and Josep Malvehy[2]

[1] Grup de Recerca en Sistemes Intel.ligents
La Salle - Universitat Ramon Llull
Quatre Camins, 2 - 08022 Barcelona
{rnicolas,dave,elisabet,afornells}@salle.url.edu

[2] Melanoma Unit, Dermatology Department
Hospital Clinic i Provincial de Barcelona, IDIBAPS
{spuig, jmalvehy}@clinic.ub.es

Abstract. Nowadays solar exposure habits have caused an important increase of melanoma cancer in the last few years. Mortality rates caused by this illness are the most important ones in dermatological cancers. Despite of it, recent studies demonstrate that early diagnosis improves drastically life expectancy. This work introduces a way to combine different kinds of diagnostic techniques to help experts in early detection. The approach is an improvement of a previous work that combines information of two of the most important non-invasive image techniques: Reflectance Confocal Microscopy and Dermatoscopy. Current work beats the results of the previous system by the support of a set of rules obtained from data preprocessing. This improvement increases the reliability of diagnosis.

1 Introduction

Sun rays and its artificial substitutes are exceedingly appreciated in our society. Actually, they could be healthy with an appropriate protection against excessive tanning. In spite of it, new social habits in solar exposure have increased the appearance of melanoma and other skin cancers (according to the American Cancer Society data). This information and the fact that melanoma holds the highest percentage of death faced with more common dermatological cancers, approximately twenty percent of non early prognosticated cases, makes even more important the early diagnosis. In order to deal with this problem, the most important way is to use non-invasive techniques based on images. In this field stands out two kinds of image analysis: Dermatoscopy and Reflectance Confocal Microscopy [12]. The former is based on the microscopical image created with epiluminiscence and the latter makes the image with the reflectance of a coherent laser with a cell resolution.

This work presents a computer aided system for medical experts in melanoma diagnosis. To catch this aim up we based our effort on solving the problems observed in our previous implemented solution [11]. Our prior approach for this

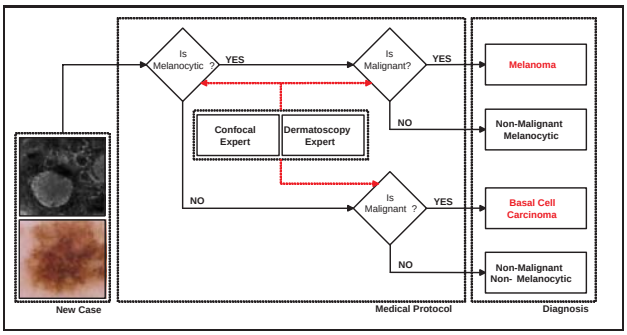

Fig. 1. Protocol followed by medical experts for melanoma diagnosis.

problem focused on the importance of the work in combination of different diagnostic criteria to ascertain the stand out of confocal in the medical protocol. We propose the use of Case-Based Reasoning (CBR) [1] techniques in order to assist the diagnosis. The CBR is a suitable approach because it uses past experiences to solve new cases. It is exactly the same procedure used by experts. Then, we use two independent CBR systems with different types of information in each one in order to obtain a shared prognostic. Moreover, we follow the medical protocol in this kind of decisions which is: to analyze whether the new case is melanocytic and, afterwards, to assess about its malignancy. Thus, the combination of both diagnosis allows experts to determine if the new case is Melanoma, Basal Cell Carcinoma (BCC) or a non-malignant tumor as figure 1 shows. With this first solution we denote that due to its high precision, Confocal microscope allows medical experts to improve its prognostic capacity making up its high economical and temporal cost and the combination of this technique with dermatoscopy allows even a better diagnostic. But this work stresses one difficulty which is that the available attributes are discrete. Considering this characteristic of the data, in our current proposal we try to tackle this property using a preprocessing technique of the data domain that is independent from the medical experts. Eventually, both techniques are combined to create a computer aided system that uses the two CBR modules to classify new cases and fetching obtained rules to combine the classification results. The whole process follows the medical protocol in this kind of decisions and the introduction of the rules guarantees more reliability in the integrated system because all rarities of the data are detected, as we will explain in following sections.

The paper is organized as follows. Section 2 describes some related work. Section 3 describes the new tool obtained by the combination of the different modules. In section 4, experimentation is presented and the performance is analyzed. Finally, section 5 summarizes the conclusions and further work.

2 Related Work

In this work, we would like to use the combination of different decisions in order to classify new melanoma cases. Although there are works focused on studying

the melanoma domain from individual approaches such as in [5], the application of ensemble methods and combination systems has increased in the last years. A line is to improve clustering using ensembles [3]. There are also works to allow the classification using data of different complexity [2] and with different types of medical information [6]. In contrast to these approaches we would like to classify melanoma following the medical diagnosis protocol using different CBR independent classifiers.

In our framework we have two available different points of view (confocal and dermatoscopic) and we select the best classification (the one prognosticated for one of the systems) depending on different criteria. These characteristics do not allow to talk about an Ensemble Learning system but the combination idea is pretty similar to the one used in this kind of methods. Ensemble methods combine the decisions from different systems to build a more reliable solution using the individual ones [9].The combination of approaches can be summarized [4] in: 1) Bagging, 2) Boosting, and 3) Stacking. Bagging and Boosting are based on the combination of the outputs using votes. Particularly, Bagging replicates N systems of the same approach with different data sources. In opposition Boosting defines complementary models. On the other hand Stacking is based on heuristics that combine the outputs of several approaches. As voting methods the most common ones [8] are: 1) Plurality, 2) Contra-Plurality, 3) Borda-Count, and 4) Plurality with Delete. All these methods are based on the number of votes of a class (plurality) but with multiple types of plurality addition and decision of better class.

Attending to the medical necessities and the existing data, it would be interesting to create a combination model with an expert for each kind of data and a final diagnosis opinion. Note that we are not using any of the classical systems of ensemble learning. As we have said, we are not formally using that technique but we are doing an adaptation of different ideas from these models. We modify the use of the data because we are using different attributes of the same data in each system, then the independence of the data is guaranteed, in contrast to the standard Bagging. Analyzing that the classification attributes are boolean, the vote method should be based on plurality but with some arranges requested by medical researchers, who weight more the information from Confocal Microscopy. Once we have experimented the protocol used by experts, we would like to improve it by using rules to do a better system combination.

In previous works in which clustering was used in order to discover new patterns on the medical domain [13] we detected a particular behavior on the data. This characteristic guarantees a correct classification of the new patients when certain conditions in the attributes of the case were detected. Thus, bearing in mind that the main goals are the improvement of the classification and to minimize the false negatives situations, we created the D-Rule module. This module preprocesses the input data and creates a set of rules to help the whole classifier. The module details are explained later on.

A similar idea is shown in [7]. In that work, the attributes of a domain are studied using overlapped intervals and computing the distance between these

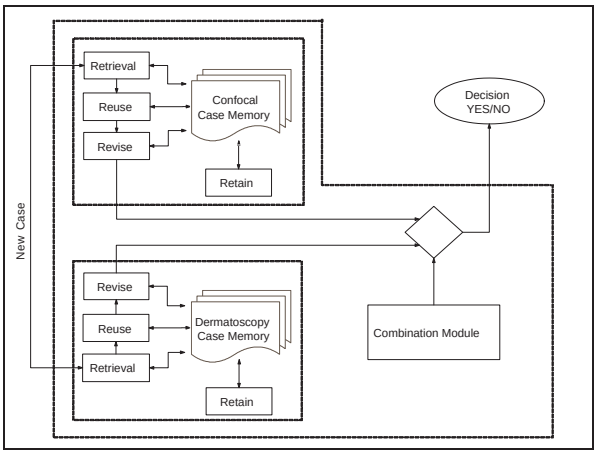

Fig. 2. Combined Decision Schema.

intervals. In this way, a technique called *overlapping of binding box* is presented and it is based on the creation of delimited intervals to define the attributes domain. However, in this paper the preprocessing module is not based on intervals. Concrete values (with useful classification) are detected and encapsulated in a rule. Moreover, our rules do not depend ones on each other, and each attribute is analyzed independently.

3 Improving a Combination of CBR Systems

In this section, we want to describe how a basic system evolves to a reliable tool for medical experts in an action framework in melanoma diagnosis. First of all, the simple combination of CBR systems is presented. Then, we introduce the D-Rule module which is capable to extract a set of characteristics of training data set using a preprocessing algorithm. Finally, we show the result of the combination of both parts.

3.1 Using the Combination of Case-Based Reasoning Systems in Dermatological Cancer

We have developed a computer aided system for melanoma diagnosis based on the medical protocol described in the first section. For each one of the decision points, a CBR system is used to answer the medical question using the knowledge extracted from the Dermatoscopy and the Reflectance Confocal Microscopy image data as Fig. 2 shows. In this first approach [11], we implement as combination module the same protocol used by medical experts.

As we can observe, the global system combines the output of two Case-Based Reasoning (CBR) systems [1]. This is because CBR performs the same resolution procedure that experts: solving new cases through the comparison of previously solved ones. In a general way, the CBR life cycle can be summarized in the next four steps: 1) Retrieving the most similar cases from the case memory with

the assistance of a similarity function; 2) Adapting the retrieved solutions to build a new solution for the new case; 3) Revising the proposed solution and 4) Retaining the useful knowledge generated in the solving process if it is necessary. Thus, the explanation capability is highly appreciated by experts because they are able to understand how decisions are made. Each one of the CBR systems feed from two different case memories which store all the previously diagnosed injuries through the confocal and the dermatoscopy studies respectively. These two parts are completely independent and at the end of its work they put on its vote for the best classification according to their specific data. With this separate ballots, the system creates the final diagnosis (Solution) and, if proceeds, saves the new case in one of the case memories or in both.

The first option tested, as a decision process to perform a diagnosis, is described in figure 3. This approach represents exactly the logical scheme used by the experts. In spite of using a collaborative scheme, where both diagnosis are combined, experts mainly focus on confocal diagnosis and, only if the diagnosis is non conclusive they use the dermatological one. Therefore, the selection of the threshold values used to perform this decision are crucial to achieve a good performance. Both values need to be defined by experts.

3.2 The D-Rule Module

One of the most important problems detected in medical environments is that the majority of attributes are discrete. This characterization of the domain produces a complicated situation when a CBR system is used in these cases. From this idea, we decide to improve our combination of classifiers with an additional tool which allows to evaluate the solution proposed in a reliable way using a set of specialized rules in the discrete attributes. So, in order to achieve this goal, the D-Rule module was designed. Analyzing the training data, a set of several rules are generated by this module. These rules summarize the data complexity in the case memory. The main goal is to represent the existing gaps in the data space with no information associated, in order to advise the classifier in this sense. On the same way, when the correct classification is guaranteed with a high reliability,

Let c_{new} be the new input case
Let $best_{confocal}$ be the most similar case using the confocal CBR
Let $best_{dermatoscopical}$ be the most similar case using the dermatoscopical CBR
Let $distance(c_i, c_j)$ be the distance between two cases c_i and c_j performed by the normalized Euclidean distance
Let $threshold_{confocal}$ be the minimal value to accept two cases as similar from the confocal point of view
Let $threshold_{dermatoscopical}$ be the minimal value to accept two cases as similar from the dermatoscopical point of view
Let *class(c)* be the class of the case c
if $distance(c_{new}, best_{confocal}) < threshold_{confocal}$ **then**
  return $class(best_{confocal})$
else
  if $distance(c_{new}, best_{dermatoscopical}) < threshold_{dermatoscopical}$ **then**
    return $class(best_{dermatoscopical})$
  return $class(best_{confocal})$

Fig. 3. Algorithm to diagnose a new case using the confocal and dermatoscopical criteria with plain combination.

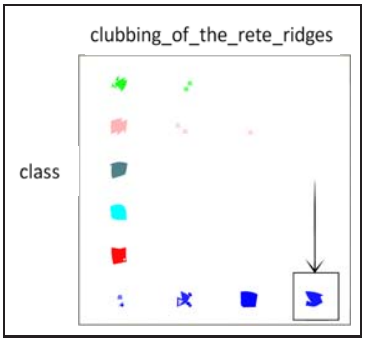

Fig. 4. Example of an attribute considered by the D-Rule module.

an alert to the following classifiers is sent, if a set of characteristics in the input data is detected.

In the example of Fig.4, we can observe a situation where a new rule is generated by the system. Looking at the *clubbing_of_the_rete_ridges* attribute, we notice that a discrete value is defined with all the cases belonging always to the same class. In this case, a new rule could be generated in order to help the CBR systems.

In our domain, it is quite usual that from a certain discrete value of an attribute, the variation of the final decision on the classification does not change. So, the intervals proposed in [7] have been eliminated and clear-cut zones affected always in the same way have been created. These zones are summarized in one or more rules.

The pseudocode used to generate the rules is described in Fig. 5. As we can observe, in the algorithm we use all possible values of an attribute to analyze the input data. One of the advantages found in the medical domain is that the attributes are well delimited, so it is not possible to find a new case with a different value of the predefined ones in the domain.

On the D-Rule module output, a set of $if - then - else$ rules are created to improve the CBR classifiers (see fig 6). The number of defined rules depends on the data complexity and it varies with the different types of classification (Melanoma, Melanocytic or BCC).

3.3 Using Rules in the Combination of CBR Systems

The final approach used in this paper is to implement the combination module based on preprocessing generated rules. Mainly, we follow the same scheme as in the first approach using the best retrieved cases but adding information

```
Let A be the set of all attributes of the medical domain
forall attributes in A do
    Let A_i be the attribute i of the set A
    Let V be the all possible values of attribute A_i
    forall values in V do
        if ∃V_j || class is unique for all cases in training set then
            CreateRule (A,i,V,j,class)
```

Fig. 5. Algorithm to generate rules in the D-Rule system.

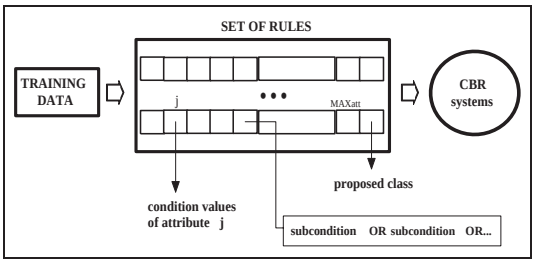

Fig. 6. Representation of the set of rules.

provided by rules. The module works with independent rules obtained from a processing technique applied to training cases. It allows the possibility of automatic and better selection of the best diagnosis. The logic process followed, as a combination, is the one described in Fig.7.

4 Experimentation

This section describes the data extracted from images and analyzes the results of the experiments performed through sensitivity and specificity rates.

4.1 Testbed

The classification of injuries in melanoma domain is not trivial. One of the main difficulties is the huge amount of information that new technologies are able to collect and the ignorance about how they are interrelated [10]. The most used techniques to gather information from tissue are the dermoscopic and the confocal analysis. Experts want to evaluate if confocal analysis detects cases impossible to assess with dermatoscope and if the usage of both techniques can improve the individual analysis.

Nevertheless, a negative point is that the confocal analysis is a long and expensive test, so the number of available cases is limited. Due to this situation,

Let c_{new} be the new input case
Let $best_{confocal}$ be the most similar case using the confocal CBR
Let $best_{dermatoscopical}$ be the most similar case using the dermatoscopical CBR
Let $distance(c_i, c_j)$ be the distance between two cases c_i and c_j performed by the normalized Euclidean distance
Let $numRules_{confocal}$ be the number of rules carried out by $best_{confocal}$
Let $numRules_{dermatoscopical}$ be the number of rules carried out by $best_{dermatoscopical}$
Let *class(c)* be the class of the case c
if $numRules_{confocal} > numRules_{dermatoscopical}$ **then**
  return $class(best_{confocal})$
else
  if $numRules_{confocal} < numRules_{dermatoscopical}$ **then**
    return $class(best_{dermatoscopical})$
  else
    if $distance(c_{new}, best_{dermatoscopical}) < distance(c_{new}, best_{confocal})$ **then**
      return $class(best_{dermatoscopical})$
    else
      return $class(best_{confocal})$

Fig. 7. Algorithm to diagnose a new case using the confocal and dermatoscopical criteria and combining the results by rules.

Table 1. Classification accuracy using only confocal images, only dermoscopic images, both images with plain combination, and both images with rules combination.

	Melanoma	Melanocytic	BCC
Only Confocal Images	87%	90%	96%
Only Dermatoscopy Images	90%	98%	95%
Confocal and Dermoscopic Images with Plain Combination	89%	96%	95%
Confocal and Dermoscopic Images with Rules Combination	94%	99%	99%

the data set used in this work is composed only by 150 instances of suspicious lesions. All instances contain information related to confocal and dermatoscopic images and the histology corroborated diagnosis. Attending to the considerations of the medical experts that have created this set, it includes enough cases from each kind of illness to be representative of the domain. Then, in medical terms it is an appropriated case memory for this study. Detailing the instances, dermatological information has forty-one fields and confocal microscopy, due to its higher resolution, contributes with data from eighty-three different attributes.

4.2 Experimentation Framework

We have tested the classification accuracy of the platform proposed with a basic decision combination and with the use of rules obtained through the use of preprocessing algorithms (as has been explained). In addition, we tested the accuracy of the two independent CBR systems (one for confocal data and another for dermatoscopy). This information is complemented with the analysis of the sensitivity and the specificity for the different cases. In the case of the plain combination platform, we have tested different decision thresholds according to medical experts criteria. The final consensus is to use 0.5 as confocal threshold and double of the distance between new case and best confocal case as dermatological one. All the CBR systems used in experimentation are configured with one-nearest neighbor algorithm with normalized Euclidean distance as retrieve function. This experiment framework has been tested applying a leave-one-out to the original data to obtain the average accuracy of those systems.

4.3 Results and Discussion

Analyzing these results, table 1 shows the accuracy rate to classify new injuries. This analysis is done in the three possible classes (Melanoma, Melanocytic and BCC) and using only confocal image, only dermoscopic image and both images with and without rules. The results obtained highlight two points: firstly, the combination of systems (even without rules) obtain a significatively better classification results, or at least an equivalent one (tested with t-test at 95% confidence level) than the use of only one kind of data. Secondly, the results are even better if we apply the preprocessing obtained rules to the combination module. In this way, we could see that the use of rules improve a minimum of 3% the accuracy in comparison with the plain combination. In all kind of classifications

Table 2. Sensitivity results using only confocal images, only dermoscopic images, both images with plain combination, and both images with rules combination.

	Melanoma	Melanocytic	BCC
Only Confocal Images	73%	94%	81%
Only Dermatoscopy Images	73%	99%	92%
Confocal and Dermoscopic Images with Plain Combination	70%	96%	92%
Confocal and Dermoscopic Images with Rules Combination	81%	100%	92%

Table 3. Specificity results using only confocal images, only dermoscopic images, both images with plain combination, and both images with rules combination.

	Melanoma	Melanocytic	BCC
Only Confocal Images	92%	81%	99%
Only Dermatoscopy Images	96%	98%	95%
Confocal and Dermoscopic Images with Plain Combination	95%	96%	96%
Confocal and Dermoscopic Images with Rules Combination	98%	98%	100%

analyzed, the increase of accuracy using rules is significant.On the other hand, tables 2 and 3 summarize the results of analyzing the statistics from the point of view of specificity and sensitivity. They show that is more reliable to do a prognostic of real negative cases than the positive ones. This happens because data sets are unbalanced, what means that, they have different number of cases of each type because data sets represent a real situation: there are more healthy people than sick people. Despite of it, the use of the combination of both types of images with the help of preprocessing obtained rules allows an important increase of sensitivity and specificity rates (in addition to improve the accuracy). It is important to highlight that, in melanoma classification, we reduce from 10 to 2 the false negatives cases and from 10 to 7 in the case of false positives. It is the most important result that medical experts would like to obtain in this kind of classification. In addition, in Basal Cell Carcinoma classification we reduce a 60% the rate of false negatives and a 100% in the case of melanocytic class. This decrease of the false negative cases (even the decrease of false positives produced too in all cases) using the combination with rules is extremely interesting and valued by medical experts.

5 Conclusions and Further Work

Melanoma early diagnosis is one of the main goals in dermatology. The diagnosis process with non-invasive techniques is complex because of the data typology. We propose a platform for automatizing the medical protocol followed in order to diagnose dermatological cancer. The proposal combines information from two promising techniques based on images through a combination algorithm based on experts' experiences. In addition, we use preprocessed rules in order to improve this combination. After the results analysis of testing melanoma data set, we can

conclude that the combination of both images improves the individual results and that by using rules we obtain even better accuracy.

The further work is focused on two lines. The first one is the improvement of the systems combination with new types of rules or other combination techniques. Secondly, the use of preprocessed rules to extract useful diagnostic patterns for medical experts.

Acknowledgments

We would like to thank the Spanish Government for supporting the MID-CBR project under grant TIN2006-15140-C03, the Generalitat de Catalunya for the support under grants 2005SGR-302 and 2008FI_B 00499, and La Salle for supporting our research group. The work performed by S. Puig and J. Malvehy is partially supported by: FIS, under grant 0019/03 and 06/0265; Network of Excellence, 018702 GenoMel from CE.

References

1. A. Aamodt and E. Plaza. Case-based reasoning: foundational issues, methodological variations, and system approaches. *AI Communications*, 7(1):39–59, 1994.
2. R. Abdel-Aal. Abductive network committees for improved classification of medical data. In *Methods of Information in Medicine*, 2004.
3. R. Avogadri and G. Valentini. Fuzzy ensemble clustering for DNA microarray data analysis. In *Lecture Notes in Computer Science 4578*, 2007.
4. E. Bauer and R. Kohavi. An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Machine Learning*, 36(1-2):105–139, 1999.
5. A. Fornells, E. Armengol, E. Golobardes, S. Puig, and J. Malvehy. Experiences using clustering and generalizations for knowledge discovery in melanomas domain. In *Advances in Data Mining. Medical Applications,E-Commerce, Marketing, and Theoretical Aspects*, volume 5077, pages 57–71. Springer, 2008.
6. A. Fornells, E. Golobardes, E. Bernadó, and J. M. Bonmatí. Decision support system for breast cancer diagnosis by a meta-learning approach based on grammar evolution. In *Eighth International Conference on EIS*, pages 222–229, 2006.
7. T. Ho, M. Basu, and M. Law. Measures of complexity in classification problems. In *Data Complexity in Pattern Recognition, pp. 1-23, Springer*, 2006.
8. K. T. Leung and D. S. Parker. Empirical comparisons of various voting methods in bagging. In *Ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 595–600. ACM Press, 2003.
9. P. Melville and R. J. Mooney. Diverse ensembles for active learning. In *ICML 04: Twenty-first international conference on Machine learning*, 2004.
10. R. Nicolas, E. Golobardes, A. Fornells, S. Puig, C. Carrera, and J. Malvehy. Identification of relevant knowledge for characterizing the melanoma domain. In *Advances in Soft Computing*, volume 49 2009, pages 55–59. Springer Berlin-Heidelberg, 2008.
11. R. Nicolas, E. Golobardes, A. Fornells, S. Segura, S. Puig, C. Carrera, J. Palou, and J. Malvehy. Using ensemble-based reasoning to help experts in melanoma diagnosis. In *Frontiers in AI and App., vol 184*, pages 178–185. IOS Press, 2008.
12. A. Scope and al. In vivo reflectance confocal microscopy imaging of melanocytic skin lesions. *J Am Acad Dermatol*, 57:644–658, 2007.
13. D. Vernet, R. Nicolas, E. Golobardes, A. Fornells, C. Garriga, S. Puig, and J. Malvehy. Pattern Discovery in Melanoma Domain Using Partitional Clustering. In *Frontiers in AI and App., 184, IOSPress*, pages 323–330, 2008.

Melanoma diagnosis based on collaborative multi-label reasoning. *Setzè Congrés Internacional de l'Associació Catalana d'Intel·ligència Artificial*, 2013

Melanoma Diagnosis based on Collaborative Multi-Label Reasoning

Nicolas RUBEN [a,1], Fornells ALBERT [a], Golobardes ELISABET [a], Corral GUIOMAR [a], Puig SUSANA [b] and Malvehy JOSEP [b]

[a] *Grup de Recerca en Sistemes Intel·ligents*
La Salle - Universitat Ramon Llull
Quatre Camins 2 - 08022 Barcelona
[b] *Melanoma Unit, Dermatology Department*
Hospital Clinic i Provincial de Barcelona, IDIBAPS

Abstract. The number of melanoma cancer-related death has increased over the last few years due to the new solar habits. Early diagnosis has become the best prevention method. This work presents a melanoma diagnosis architecture based on the collaboration of several multi-label case-based reasoning subsystems called DERMA. The system has to face several challenges that include data characterization, pattern matching, reliable diagnosis and self-explanation capabilities. Experiments using two subsystems specialized in confocal and dermoscopy data from images respectively have provided promising results to help experts assess melanoma patterns.

Keywords. Melanoma Cancer Diagnosis, Case-Based Reasoning, Collaborative Systems, Multi-Label, Distance Metric Learning

Introduction

Melanoma has become one of the most relevant social cancers because it is becoming more prevalent in our society and it affects people of any age. Although it is not the most common skin cancer, its mortality is around twenty percent if it is not treated early, according to the American Cancer Society. This situation has fomented the application of artificial intelligence techniques to exploit data and help experts in the early diagnosis.

This work describes DERMA, an architecture created as a result of the collaboration with the department of dermatology at the Hospital Clinic of Barcelona (HCPB) and the Institute of Biomedical Research August Pi i Sunyer (IDIBAPS). DERMA diagnoses thanks to a collaborative architecture of several subsystems which specialize in different kinds of databases. More specifically, the current version is based on the collaboration [4] of two multi-label Case-Based Reasoning (CBR) [1] systems that use confocal and dermoscopy data respectively. CBR is used as engine due to its self-explanation capabilities extracted from solving new problems using past experiences, which are of

[1]Corresponding Author: Ruben Nicolas; E-mail: rnicolas@salle.url.edu.

great importance to experts when it comes to understanding results. Moreover, the CBR data is reorganized using a distance learning metric [25] to improve the identification of useful knowledge and the CBR classification is extended to work in multi-label mode [2]. Multi-label means that a new case can be classified in more than one class so this additional information consequently helps experts to understand best complex patterns because they are be represented as a set of simple patterns. For this reason, distance metric learning and the multi-label collaboration scheme have led to improvements in the sensitivity and specificity of diagnosis.

The following sections describe DERMA in detail. Section 1 presents the DERMA architecture and its modules. Next, Section 2 highlights the main results obtained. Finally, Section 3 ends with the conclusions and further work.

1. DERMA: Melanoma Diagnosis based on Collaborative Multi-Label Reasoning

DERMA is a platform that aids medical experts in melanoma diagnosis. Figure 1 describes the DERMA architecture which addresses four different challenges identified during the collaboration with HCPB that range from data acquisition to diagnosis. The first challenge focuses on the creation of a melanoma ontology [11] based on a characterization of the domain performed through interviews with melanoma experts and the study of melanoma patterns. The second one is to create specialized subsystems in order to work with the different data sources. CBR [1] is selected due to its suitability for working in environments where self-explanations are required. This step also considers how to properly organize the knowledge bases in order to improve its performance through distance metric learning [25]. The third challenge is to define a collaborative scheme [4] between the independent CBR subsystems based on the modus operandi of the experts and also to include mechanisms to manage exceptional situations. Finally, the last challenge is related to the complex task of making classifications of non trivial patterns. In

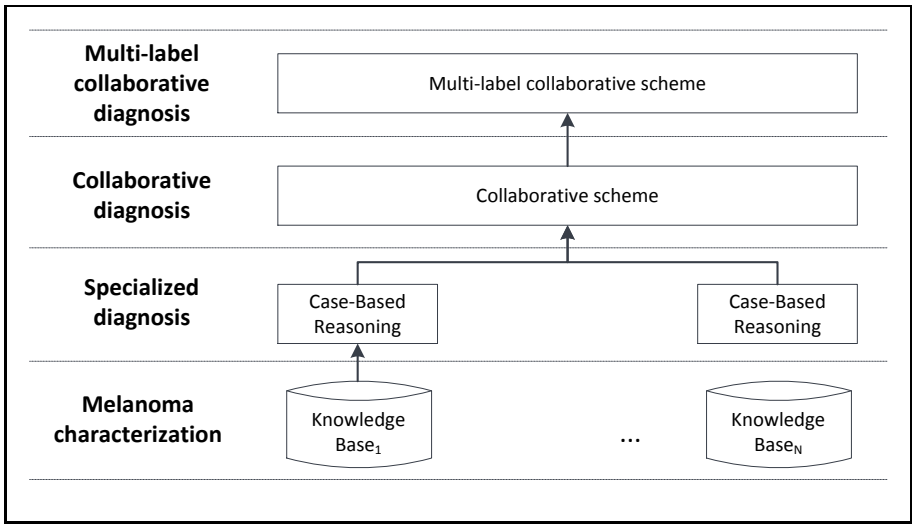

Figure 1. Melanoma diagnosis architecture based on collaborative multi-label reasoning.

this sense, DERMA may be able to learn and diagnose better if non trivial patterns could be represented as a set of simple patterns. For this reason, we decided to migrate the single-label CBR subsystems to multi-label [2] CBR subsystems. The next subsections describe these challenges in detail.

1.1. Challenge 1: Melanoma Characterization

The first step in the development of any knowledge-based system is to identify, understand and gather data associated to the problem. These steps are non trivial when the system is related to the health sciences domain because these data are usually characterized as being heterogenous and coming from several medical profiles involved in the prognosis. Moreover, experts often label data according to their interests and background so they may use different names to refer to conditions which have the same meaning. Thus, the characterization and understanding of the relationships between all data sources in order to plan or gather knowledge is a non trivial task that requires time and consensus between experts.

The specific domain characterization was performed using data from more than three thousand patients with melanoma cancer and contained reports from dermatologists, oncologists, surgeons, pathologists and other specialists working in HCPB. As in the great majority of medical problems, data was heterogenous and distributed in different plain databases and many attributes were represented and stored differently according to the expert. For this reason, an ontology [11] was defined using the experts' point of views

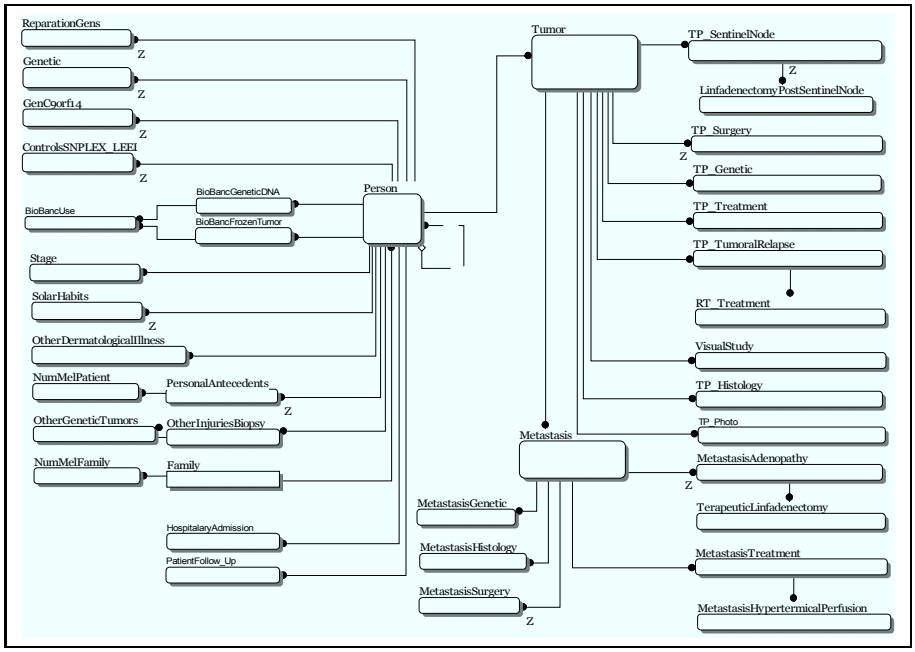

Figure 2. The melanoma relational model allowed the integration of data sources from the different medical profiles. The model considers data of the patient, its family, generic information, tumors, metastasis, controls, and studies.

and data from international studies that examine specific aspects of the domain [14]. Although this unified point of view permitted the integration between all data sources as Figure 2 shows, there were pieces of information that could not be integrated because experts did not have records regarding to all the patients. For this reason, we decided to focus on the usage of nevus images analysis due its availability. More specifically, Dermoscopy and Confocal images were selected because they are two of the most promising image analysis techniques of image analysis for the diagnosis of melanoma. Dermoscopy is based on a microscopic image created epiluminiscence (x10.30) and confocal reflectance is generated by the reflection of a coherent laser (x100) resolution at the level of the cell.

Finally, we wanted to know if would be possible to identify equivalent patterns by using the same data used by medical experts. Melanoma diagnosis is based mainly on the ABCD rule which considers the following characteristics typically observed in this type of tumor: (A) a diameter greater than 5 mm, (B) color variation, (C) asymmetry, and (D) jagged edges. We tested K-means [10] and Self-Organizing Maps (SOM) [12] due to our previous experiences in breast cancer diagnosis [7]. K-means makes a partition of the domain in K clusters and SOM translates complex and non-linear statistical relations contained in high-dimensional data into simple geometric relations on a low-dimensional space which provide an optimal organization. The results provided equivalent patterns to the ones identified by ABCD rules [23,5].

1.2. Challenge 2: Diagnosis using the useful knowledge

CBR [1] systems solve new problems through an analogical procedure based on experiences represented by a set of cases stored in a case memory. Thus, CBR is able to justify the obtained solutions using analogies with previous problems which is crucial for experts. The way in which CBR works can be summarized in the following steps: (1) it retrieves as many similar cases as possible from the case memory with a similarity function, (2) it adapts them to propose a new solution, (3) it checks if this solution is valid, and finally, (4) it retains the useful information of the prognostic if it is necessary. All CBR steps turn around the case memory round and so its organization and the way cases are retrieved determine its performance in terms of accuracy[2] and computational time [8].

There are two main possible memory organizations: flat and structured. A flat organization is the simplest way because cases are stored sequentially in a list. In this situation, the strategy for classifying a new problem is to sequentially compare it with all the cases in that list using some measure of similarity. The main shortcoming of this approach is that the more cases the case base contains, the higher the time of retrieval. As for the structured memory organization, many authors have tackled this issue from many points of view, such as representing the attributes in tree structures [24] or graphs [13], grouping cases by their similarity [26], applying knowledge-intensive approaches [20] or data-intensive approaches [9]. On the other hand, the ideal similarity function definition depends on the domain and how data is related. There are even works which try to discover this similarity function using algorithms based on genetic algorithms [6], the application of standard similarity functions such as Euclidean distance are the most often solution because they are more reliable.

[2]We understand accuracy as the capability of minimizing the sensitivity and specificity

Due to the importance of obtaining an optimal determination of positive and negative melanoma cases we propose a data organization based on Distance Metric Learning (DML) [25] as a technique to identify a suitable distance metric based on the data projection. DML uses a metric that keeps all the data points from the same class together and, at the same time, separates the data points from different classes as far as possible. We learned a global distance metric that minimized the distance between pairs of data included in the equivalence constraints and data pairs from the inequivalence constraints [18]. With this process we obtained a case memory which was organized in a way that permitted a better retrieval given the fact that the positive and negative cases become distanced.

1.3. Challenge 3: Improving the reliability through collaborative schemes

The way in which melanoma is diagnosed takes into account different data sources which make it suitable for working with a collaborative approach [3]. Figure 3 describes the medical process integrated in DERMA [19]. The combination through the medical protocol determines if the new case is melanoma, Basal Cell Carcinoma (BCC), or a non-malignant tumor (melanocytic or not) according to the single diagnosis provided by the specialized CBR subsystems using the confocal and dermoscopy image of the new patient. The analysis of the initial results showed the supremacy of the confocal study in medical protocol. The high precision confocal microscope enables medical experts to improve their ability to forecast, thus justifying its high cost and time. Furthermore, the combination of this technique with dermoscopy leads to an even better diagnosis. Additionally, a knowledge rules module is introduced to ensure the reliability in exceptional situations due to data oddities. The idea is to weight the single classification of each subsystem according to the reliability of the retrieved cases. The reliability of a case is based on a set of rules previously extracted from data. This new step adds a fine tune to the classifiers collaboration that improves the final classification. Finally, a test with all previous modules was done and the experimentation achieved the desired results as section 2 shows [18].

1.4. Challenge 4: Multi-Label Diagnosis

During the study of the characteristics of the medical protocol we saw that melanoma problem is better fitted by a multi-label classification. We reached the conclusion that is better to consider two non-disjunctive classes such as melanocytic and malignant instead of using just a melanoma class. Categorization issues (emotions, text, time, ...) or verification of faces at airports among others, are better represented with multi-label datasets than with a single label because non trivial patterns are easier to represent using a set of trivial patterns requires the proposed methods to deal with this type of data. Existing works to date have been divided into two distinct families: (1) to adapt the data set to work with single label algorithms instead of designing new algorithms. This group of techniques are known as Problem Transformation Methods (PTM) and the main problem is that the unification of the different labels may result in the loss of information, (2) to address the problem with a modification of the classical algorithms to work purely multi-label which is known as Algorithm Adaptation Methods (AAM) [21].

The reuse phase of the CBR subsystems was extended to work with multi-label using the AAM approach because we did not want to lose any information during the clas-

sification process [17]. First of all we proposed an adaptation of the reuse to multi-label classification in a probabilistic manner. It is based on the idea of counting the occurrence of each label and giving it the value of 1 if more than a half of the k retrieved cases have the value set to 1. This is based on the idea of the multi-label k-nearest neighbor algorithm. Later on, and in order to improve the accuracy, we tune the retain phase complementing the probabilistic reuse with experience. With this step we weight the cases by considering the results of previous classifications. We saw that this second option improves the classification results in some groups of data and there are no significant differences in other cases.

2. Experiences, Experiments, and Results

The active collaboration with medical experts has played an important role in carrying out the different stages designed in conjunction with medical experts according to their requirements. Although dermoscopy and confocal are the main data sources used by our system, not all the possible information designed in ontology has been used due to the fact it was not available. Given this situation, the data set used in this work is composed of 150 cases of suspicious lesions. For all theses cases we have information related to confocal and dermatoscopic images and the histology that corroborated diagnosis. In response to the considerations of the medical experts that have created this set, it includes enough cases from each kind of illness to be representative of the domain. Then, in medical terms it is an appropriated case memory for this study. Detailing the instances, dermoscopy information has forty-one fields and confocal microscopy, due to its higher resolution, contributes with data from eighty-three different attributes. Improvements

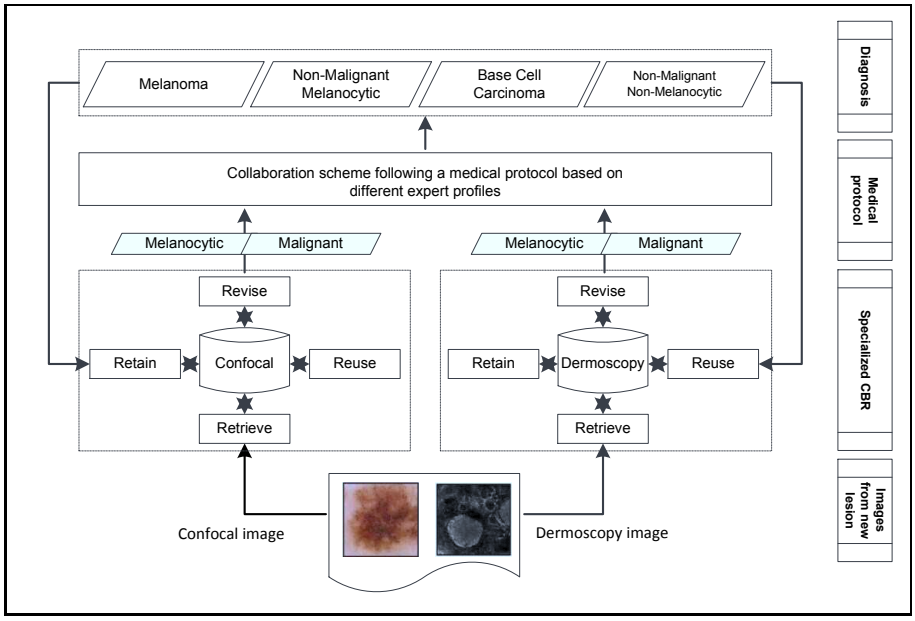

Figure 3. Collaborative scheme between specialized CBR subsystems for melanoma cancer diagnosis.

such as multi-label CBR has been tested using other synthetically generated datasets to see the global strength of the method and deal with the lack of an available multi-label melanoma dataset.

The next points summarize the most important milestones achieved and the outstanding results:

- Melanoma characterization: During this process we obtained two outstanding results: (1) an ontology for characterizing melanoma that takes into account the different medical profiles involved in diagnosis where the melanoma concept was divided into the following main groups: person/family, generic information, tumors, metastasis, controls/studies [14,15]; and (2) a set of new clusters for data classification [23], different to the ones used to the date, that medical experts considered interesting due to its fit to real cases. This process was tested using the medical experts expertise because it is a formal vision of their knowledge and a characterization of different possible classifications.
- Specialized diagnosis: This step consolidates the need of a classification process and determines the organization of the data. According to the characteristics of the domain and the medical needs we propose the integration of distance metric learning in the CBR subsystems [18] because it improves the retrieval of good cases as it increases the distance between cases of different nature. With this change in the regular classification we meet the goals of the medical experts in this area.
- Collaborative diagnosis: This challenge shows us that using a basic collaborative CBR system (MEDIBE [16]) the classification results are, in terms of sensitivity, significantly better than the ones obtained without using a collaborative scheme, which is what experts are looking for. In addition, if we apply pre-processing rules to the collaborative process (RB-MEDIBE [19]) we obtain results that are better or equivalent to the ones obtained by the non-ruled system.
- Multi-label diagnosis: Melanoma cancer is a multi-label problem as we saw when we studied its characteristics, as a consequence we moved our efforts to a multi-label classification. As we do not have any available multi-label melanoma database we have used some multi-label benchmarks and synthetically generated datasets instead of the ones used with the other parts of the platform.The results obtained with both types of data show that the multi-label approach (MlCBR [17]) should increase (or at least not decrease) the accuracy of the more competent platforms on multi-label classification, such as RAkEL or MlKnn [22,27].

Finally, we would like to graphically highlight the comparative results obtained by the different steps of DERMA. The comparison will be done between Confocal CBR (non-collaborative CBR classification using just confocal data), Dermoscopy CBR (non-collaborative CBR classification using just dermoscopy data), MEDIBE (plain collaborative system), RB-MEDIBE (collaborative system that enhance the collaboration with pre-processing rules), and COMEDI-CBR (collaborative system with a DML organized case memory). The results obtained using the multi-label version of the system (MlCBR) are not presented in the figure because this part of the platform has been tested with non-melanoma datasets due to the absence of enough multi-label melanoma data. As we can see in Figure 4 the different challenges fitted by DERMA system improve the results of previous works in terms of Accuracy, Sensitivity, and Specificity. Despite the original

rates of false negatives being high, this is a result that medical experts want to improve as sensitivity is cruciality to medical diagnosis. This goal has been accomplished with our system. Although we achieve hundred percent accuracy we are not facing a foolproof system but one which knows how to work with the peculiarities of the domain, just as medical experts. Once we obtain more data we must retest the system to see if follows the same trend.

3. Conclusions and Further Work

DERMA was proposed to work as a Decision Support Systems to help medical experts in their diagnosis. The general architecture allows the integration of several data sources that correspond to the different medical profiles and techniques involved in a melanoma diagnosis. Each data source is used to configure a single CBR system that performs a single classification. The combination between all CBR subsystems through a medical protocol scheme provides us with a final collaborative diagnosis.

Every challenge we have addressed has improved the results of previous steps. We started our work with a single classification, using a single CBR system with just one kind of data. This first step allows us to determine the base accuracy. Later we propose a collaborative diagnosis, where different subsystems make a unified prognostic with different data sources. This module performs the same process used by medical experts combining different criteria. The basic collaboration was followed by the application of enhancements on the knowledge base organization and the collaborative process. This works tune the system in order to improve the sensitivity and specificity rates. The final

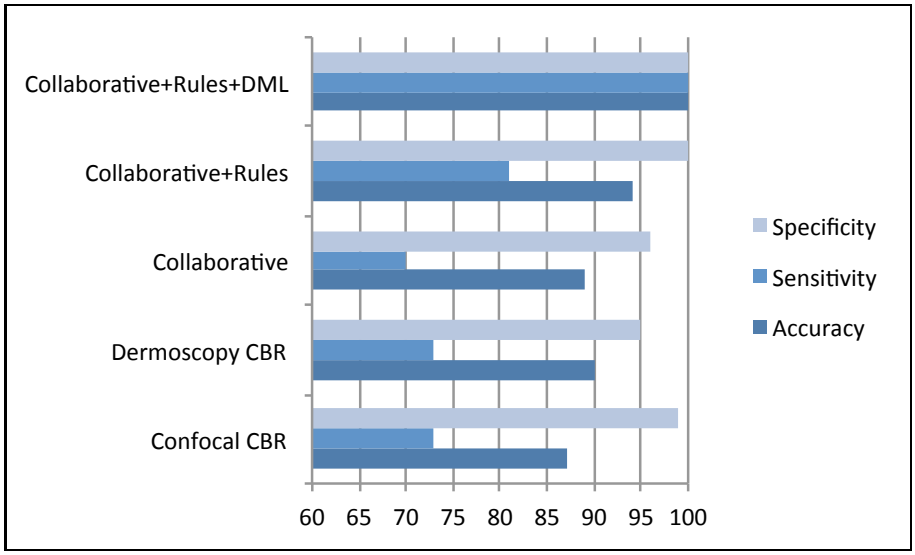

Figure 4. Accuracy, Sensitivity and Specificity results obtained through the different DERMA's challenges: the non-collaborative CBR classification using just confocal data (Confocal CBR), the non-collaborative CBR classification using just dermoscopy data (Dermoscopy CBR), the plain collaborative system (MEDIBE), the collaborative system that enhance the collaboration with pre-processing rules (RB-MEDIBE), and the collaborative system with a DML organized case memory (COMEDI-CBR).

stage involved an extension of the work to the use of multi-label data due to the domain characteristics. This last step offers good results and keeps the door open to the creation of richer datasets. We could summarize that the results obtained by DERMA during the test process were the ones expected by medical experts. Nowadays the most outstanding problem remains the lack of melanoma data but we are developing a data management application that will solve this problem. Once we obtain the complete data that fits all the melanoma features we will re-check our results in order to use DERMA system to aid the day to day assistance which is our main future goal.

On the other hand, analog reasoning and hybrid systems, such as collaborative and multi-label CBR, are hot research topics and especially interesting in the health sector. This is because it is crucial to have a flexible framework like the ones offered by this family of techniques and the system must be highly reliable within the community of data mining. This is why some of these techniques are being explored in new fields such as social networks and marketing areas. Techniques such as analog reasoning exploit the high capacity of the computer to find patterns that lead to new and useful information for the user. Our proposals could be moved to these new areas in order to take care of new problems. This is the second part of our further work.

Acknowledgements

We thank the Agencia de Gestio d'Ajuts Universitaris i de Recerca for funding the Grup de Recerca en Sistemes Intel·ligents as recognized group (2009-SGR-183), and for grant 2010FI_B2 0084.

References

[1] A. Aamodt and E. Plaza. Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. *AI Communications*, 7:39–59, 1994.

[2] J. Avila, E. Gibaja, and S. Ventura. *Multi-label Classification with Gene Expression Programming*, volume 5572 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 2009.

[3] E. Bauer and R. Kohavi. An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Machine Learning*, 36(1-2):105–139, 1999.

[4] T. G. Dietterich. Ensemble learning. In *The Handbook of Brain Theory and Neural Networks*, pages 405–408. The MIT Press, 2002.

[5] A. Fornells, E. Armengol, E. Golobardes, S. Puig, and J. Malvehy. Experiences using clustering and generalizations for knowledge discovery in melanomas domain. In *Advances in Data Mining. Medical Applications,E-Commerce, Marketing, and Theoretical Aspects*, volume 5077, pages 57–71. Springer, 2008.

[6] A. Fornells, J. Camps, E. Golobardes, and J. Garrell. Comparison of strategies based on evolutionary computation for the design of similarity functions. In *Artificial Intelligence Research and Development*, volume 131, pages 231–238. IOS Press, 2005.

[7] A. Fornells, E. Golobardes, J. Martorell, and J. Garrell. Patterns out of cases using kohonen maps in breast cancer diagnosis. *International Journal of Neural Systems*, 18:33–43, 2008.

[8] A. Fornells, E. Golobardes, J. Martorell, J. Garrell, E. Bernado, and N. Macia. A methodology for analyzing the case retrieval from a clustered case memory. In *7th International Conference on Case-Based Reasoning*, volume 4626 of *LNAI*, pages 122–136. Springer-Verlag, 2007. Best paper nomination.

[9] A. Fornells, E. Golobardes, D. Vernet, and G. Corral. Unsupervised case memory organization: Analysing computational time and soft computing capabilities. In *8th European Conference on Case-Based Reasoning*, volume 4106 of *LNAI*, pages 241–255. Springer-Verlag, 2006.

[10] J. Hartigan and M. Wong. A k-means clustering algorithm. In *Applied Statistics*, pages 28:100–108. 1979.

[11] Y. Kalfoglou and M. Schorlemmer. Ontology mapping: The state of the art. *The Knowledge Engineering Review*, 18:1–31, 2003.

[12] T. Kohonen. *Self-Organizing Maps*. Springer, 3rd edition, 2000.

[13] M. Lenz, H. Burkhard, and S. Brückner. Applying case retrieval nets to diagnostic tasks in technical domains. In *Proc. of the 3rd European Workshop on Advances in Case-Based Reasoning*, pages 219–233. Springer-Verlag, 1996.

[14] R. Nicolas, E. Golobardes, A. Fornells, S. Puig, C. Carrera, and J. Malvehy. Identification of relevant knowledge for characterizing the melanoma domain. In *Advances in Soft Computing*, volume 49 2009, pages 55–59. Springer Berlin-Heidelberg, 2008.

[15] R. Nicolas, E. Golobardes, A. Fornells, S. Puig, and J. Malvehy. Estudi de les caracteristiques del domini del melanoma (code: 071208). Technical report, EALS-URL, 2008.

[16] R. Nicolas, E. Golobardes, A. Fornells, S. Segura, S. Puig, C. Carrera, J. Palou, and J. Malvehy. Using ensemble-based reasoning to help experts in melanoma diagnosis. In *Frontiers in AI and App., vol 184*, pages 178–185. IOS Press, 2008.

[17] R. Nicolas, A. Sancho-Asensio, E. Golobardes, A. Fornells, and A. Orriols-Puig. Multi-label classification based on analog reasoning. *Expert Systems With Applications Journal*, page In Press, 2013.

[18] R. Nicolas, D. Vernet, E. Golobardes, A. Fornells, F. de la Torre, and S. Puig. Distance metric learning in a collaborative melanoma diagnosis system with case-based reasoning. In *Proceedings of the 14th United Kingdom Workshop on Case-Based Reasoning at The 29th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence, CMS Press, University of Greenwich, UK*, pages 58–66, 2009.

[19] R. Nicolas, D. Vernet, E. Golobardes, A. Fornells, S. Puig, and J. Malvehy. Improving the combination of cbr systems with preprocessing rules in melanoma domain. In *Workshop Proceedings of the 8th International Conference on Case-Based Reasoning, Seattle, WA, USA*, pages 225–234, 2009.

[20] E. Rissland and J. Daniels. The synergistic application of cbr to ir. *Artificial Intelligence Review*, 10(5):441–475, 1996.

[21] R. E. Schapire. Boostexter: a Boosting-Based System for Text Categorization. *Machine Learning*, 39:135–168, 2000.

[22] G. Tsoumakas and I. Vlahavas. Random k-labelsets An ensemble method for multilabel classification. In *Proceedings of the 18th European Conference on Machine Learning*, pages 406–417, 2007.

[23] D. Vernet, R. Nicolas, E. Golobardes, A. Fornells, C. Garriga, S. Puig, and J. Malvehy. Pattern discovery in melanoma domain using partitional clustering. In *Frontiers in AI and App., vol 184*, volume 184, pages 323–330. IOS Press, 2008.

[24] S. Wess, K. Althoff, and G. Derwand. Using k-d trees to improve the retrieval step in case-based reasoning. In *1st European Workshop on Topics in Case-Based Reasoning*, volume 837, pages 167–181. Springer-Verlag, 1994.

[25] E. P. Xing, A. Y. Ng, M. I. Jordan, and S. Russell. Distance metric learning, with application to clustering with side-information. In *Advances in Neural Information Processing Systems 15*, pages 505–512. MIT Press, 2002.

[26] Q. Yang and J. Wu. Enhancing the effectiveness of interactive case-based reasoning with clustering and decision forests. *Applied Intelligence*, 14(1), 2001.

[27] M.-L. Zhang and Z.-H. Zhou. Ml-Knn: A Lazy Learning Approach to Multi-Label Learning. In *Pattern Recognition*, volume 40, pages 2038–2048, 2007.

Intelligent tutoring system framework for the acquisition of knowledge and competences. *40th ASEE/IEEE Frontiers in Education Conference*, 2010

Session T1A

Work in Progress – Intelligent Tutoring System Framework for the Acquisition of Knowledge and Competences

David Vernet, Ruben Nicolas, Elisabet Golobardes, Albert Fornells and Alvaro Garcia-Piquer
La Salle, Universitat Ramon Llull, dave@salleurl.edu, rnicolas@salleurl.edu, elisabet@salleurl.edu, afornells@salleurl.edu, alvarog@salleurl.edu

Abstract – **Even though there are many tutoring systems to help students achieve their goals in terms of theoretical knowledge, as yet there is no system to foment the acquisition of the competences which form an integral part of university degree programs. This issue is crucial because the Higher Education System is changing in Europe. New educational models are being created to introduce competences which correspond specifically to degree programs. In this work-in-progress the general framework to develop an Intelligent Tutoring System (ITS) based on competences is presented. The system monitorizes the acquisition of theoretical knowledge and also the assessment of the competences related to the subject and it takes corrective actions when needed to fix a negative evolution of the student. The framework is divided into four main phases based on artificial intelligence techniques: construction, location, prediction and reinforcement. The main feature of the proposed framework is the fact that it promotes the academic development of the student in the future educational context by providing guidance and supervision to ensure the successful acquisition of both theoretical knowledge and the corresponding subject-related competences.**

Index Terms – Assessment, Competences, Distance Learning, Intelligent Tutoring Systems.

INTRODUCTION

The new Higher Education Area has produced a number of changes in teaching and assessment methods. [1]-[2]. In the past, the weight of the learning process was borne by the teacher and the student was a participing user of the class. In the new teaching model, the student is at the centre of the learning process and the objective is not only based on the acquisition of knowledge, but also on competences (this word refers to a set of skills, knowledge and attitudes that an individual must possess in order to be capable of perform a specified job [1]).

This new model, however, presents us with a handicap which is that most distance-learning teaching tools have not been thought out for this new framework. Practically all the systems focus on the acquisition of knowledge and ignore the acquisition of competences. Some examples are [3]-[4].

In order to cover this shortfall we propose to create an intelligent tutoring system frame for the acquisition of Knowledge and Competences. The work is still in progress but we have started to define the most relevant features of this type of system.

The following sections define the structure of a system of this nature and its development framework. The framework is divided into four phases: construction, location, prediction and reinforcement. The following sections will describe each phase in greater detail. Firstly, however, we shall present the data which the system will work with. This is specific data which has to be collected and analyzed in order to carry out an evaluation of the competences.

DATA DEFINITION

One of the most important features of the new Higher Education Area is that each subject is assigned a series of competences which have to be acquired by the student and assessed by the teacher [5]. Taking things to extreme, we could define the final assessment of a subject as being the weighted assessment of these competences. In other words, the acquisition of a competence is computed in line with the knowledge areas which evaluate the competence. The final mark does not depend on one particular topic, but on a set of knowledge areas which evaluate the competence. The results obtained correlate with the acquisition of theses competences which shape the final course mark.

Given this situation, the data used by the system is predetermined. A set of competences is assigned to each knowledge area of the subject and a subsequent set of activities is defined for each competence. This enables us to assess the progress of the student who is acquiring this competence. Figure 1 shows the distribution in terms of areas of knowledge. The letters are coded as follows:

- S = Subject, the subject to teach.
- Ui = Unit of the subject (1<=i<=N). In the traditional teaching a subject is divided in units.
- Ai = Activity associated to the unit in order to assess the learning of the student.
- Ci = Competence associated to activity. Each activity may be associated to more than one competence.

/10/$25.00 ©2010 IEEE

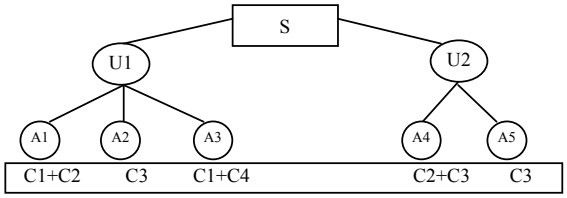

FIGURE 1
TRADITIONAL REPRESENTATION OF THE KNOWLEDGE TREE

As can be seen in Figure 1, the knowledge of the subject is represented according to the model of traditional teaching and way which the most important systems are currently using to represent it.

Our approach sets the competence model as the center of the knowledge to be taught. Figure 2 shows how the previously-mentioned knowledge tree would be transformed in the new teaching model.

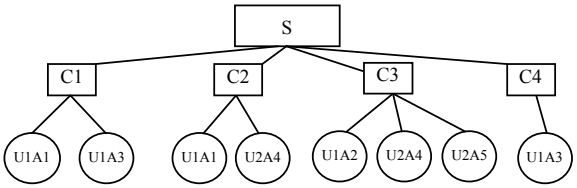

FIGURE 2
REPRESENTATION OF THE DATA IN OUR FRAMEWORK

The diagram shows how the tutoring system focuses on all of its operations on the competence and how the activities proposed throughout the learning process of the student are used to optimize the potential of each competence.

THE FRAMEWORK

The point of view used of relating the information is the operational base of our tutor system. The activities which are proposed to the students are closely linked to the competences they will obtain if they successfully complete the activity in question. The system automatically detects when a student does not acquire a specific competence and suggests new activities to enable the student to reach the required competence level. The most important data introduced into the system are the qualifications of the competences of each of the activities carried out.

The system can be divided into 4 clearly marked phases:

I. Construction

In the construction phase the knowledge tree of the subject to be taught is obtained. The nodes of this tree constitute units of the domain of knowledge which, at the same time, are divided into different activities. Unlike the *learning paths* proposed by Falmagne [6], the teachers of the subject design the activities to solve, associate competences to them and place them in the tree, depending on the competence in question. As we can see, in our system not only is theoretical knowledge placed on the tree but the competences achieved through the resolution of these activities are also added. If we observe the tree generated in the traditional way, the competences are distributed in the tree to one or several scattered nodes.

II. Location

In the location phase the system assigns the student a state of knowledge and competences using an automatic algorithm extracted from different evaluations obtained through the activities proposed by the teachers. Thus, using the information gathered from the activities already carried out, the system then assesses the degree to which every competence is reached and generates the tuple (competence-achievement grade) for each of the competence designed for the course. This evaluation and location should be repeated more than once in accordance with the level of acquired competences and knowledge in the different fields. At this point the system can activate the predictive phase.

III. Prediction

The prediction phase is based on the performance of a student (exercise marks and acquired competences), and it predicts the evolution of the student at the end of course using Artificial Intelligence techniques.

IV. Reinforcement

In the reinforcement phase an intelligent selection of suitable activities enables the system to give the student activities which not only improve learning potential of a specific subject but also develop the tree of knowledge and competences.

CONCLUSIONS AND CURRENT WORK

A new framework approach for education has been presented. The new features of it are: 1) A new way to organize the knowledge of a subject and, 2) The design of an ITS framework using this new model. Currently, we are developing an ITS with this framework and integrating AI techniques.

ACKNOWLEDGEMENTS

We would like to thank the AGAUR for the support under grants 2010FI_B2 00084 and 2010FI_B 01084, and Lisa Kinnear for the English revision.

REFERENCES

[1] Golobardes, E et al., "Guía para la evaluación de competencias en el área de Ing. y Arq.", *Guías de eval. de competencias, AQU Cat*, 2009.

[2] Hager, P, Gonczi, A, "General issues about assessment of competence, Assessment and Evaluation in Higher Education", Vol 19 (1), University of Bath, Bath, 1994.

[3] Hockemeyer, C., Held T., Albert, D., "A relational adaptive tutoring hypertext WWW-environment based on knowledge space theory", *Proceedings of CALISCE'98*, 1998, 417-423.

[4] Wetzel, M., & Hanley, G. "Evaluation of MERLOT Tools, Processes, and Accomplishments". *Center for Usb. in Design and Assess.*", 2001

[5] Bologna declaration, "The European Higher Education Area", *Joint declaration of the European Ministers of Education*, 1999.

[6] Falmagne, J.-C. "A latent trait theory via stochastic learning theory for a knowledge space", *Psychometrika*, 53, 1989, 283–303.

Distance metric learning in a collaborative melanoma diagnosis system with Case-Based Reasoning. *29th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence, United Kingdom Workshop on Case-Based Reasoning*, 2009

Applying Distance Metric Learning in a Collaborative Melanoma Diagnosis System with Case-Based Reasoning

Ruben Nicolas[1], David Vernet[1], Elisabet Golobardes[1], Albert Fornells[1], Fernando de la Torre[2], and Susana Puig[3]

[1] Grup de Recerca en Sistemes Intel.ligents
La Salle - Universitat Ramon Llull
Quatre Camins, 2 - 08022 Barcelona
{rnicolas,dave,elisabet,afornells}@salle.url.edu

[2] Robotics Institute - Carnegie Mellon University
5000 Forbes Ave., Pittsburgh, PA 15213, U.S.A.
ftorre@cs.cmu.edu

[3] Melanoma Unit, Dermatology Department
Hospital Clinic i Provincial de Barcelona, IDIBAPS
spuig@clinic.ub.es

Abstract. Current social habits in solar exposure have increased the appearance of melanoma cancer in the last few years. The highest mortality rates in dermatological cancers are caused for this illness. In spite of it, recent studies demonstrate that early diagnosis increases life expectancy. This work introduces a way to classify dermatological cancer with highest rates of accuracy, specificity and sensitivity. The approach is the result of the improvement of previous works that combine information of two of the most important non-invasive image techniques: Reflectance Confocal Microscopy and Dermatoscopy. Current work achieve better results than the previous systems by the use of Distance Metric Learning to the different Case Memories.

Keyword. Case-Based Reasoning, Distance Metric Learning, Collaborative Systems.

1 Introduction

Twenty-first century society exceedingly appreciates the exposure to sun rays and its artificial substitutes. An appropriate protection and not exceeding the recommended tanning times makes it a healthy habit. Despite of it, long solar exposure without protection has made the appearance of melanoma and other skin cancers increase. According to the American Cancer Society [4] although is not the most common skin cancer, it is the one that causes most deaths (a twenty percent of non-early diagnosed cases). To deal with this problem, the most important way is the use of non-invasive techniques based on images. In this field stand out: Dermatoscopy and Reflectance Confocal Microscopy [16]. The former is based on the microscopical image created with epiluminiscence and the latter makes the image with the reflectance of a coherent laser with a cell resolution.

With this paper we present a Collaborative Computer Aided System for melanoma diagnosis that uses Distance Metric Learning (DML) to improve its operation. To achieve this aim, we base our effort on solving the problems observed in our previous implemented solutions [14,15]. The main aim of this work is to improve the classification accuracy and, moreover, the specificity and sensitivity rates that are crucial for medical experts. In order to achieve this goal, we take a second step in our system, in this case working on the case memories of our sub-systems applying DML [19] to establish a new projection of the data that allows a better distance analysis. The DML approach we have used is the one proposed by Xing [18] using Convex Programming.

In our first attempt for this problem [14], we focused our effort on the combination of different diagnostic criteria to ascertain the stand out of confocal method in comparison with other techniques. We proposed the use of Case-Based Reasoning (CBR) [1] techniques in order to assist the diagnosis. The CBR is a suitable approach because it follows exactly the same procedure used by experts. Then, we use two independent CBR systems with different types of information in each one in order to obtain a shared prognostic. Moreover, we follow the medical protocol in this kind of decisions which is: to analyze whether the new case is melanocytic and, afterwards, to assess about its malignancy. Thus, combining both diagnosis allow experts to determine if the new case is Melanoma, Basal Cell Carcinoma (BCC) or a non-malignant tumor as figure 1 shows. With this first solution, we denote that due to its high precision, Confocal microscope allows medical experts to improve its prognostic

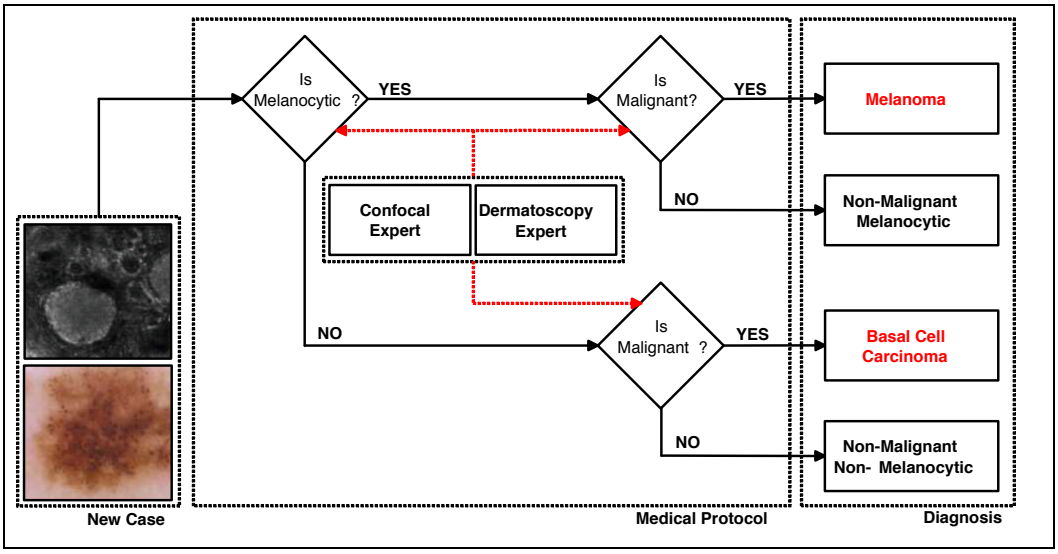

Fig. 1. Protocol followed by medical experts for melanoma diagnosis.

capacity making up its high economical and temporal cost and the combination of this technique with dermatoscopy allows even a better diagnostic. Despite this, the work stresses some difficulties that need to be solved. Considering the characteristics of the data, in our second proposal [15] we add a module that discovers rules from the data that are independent from the experts. Eventually, both techniques are combined to create a computer aided system that uses the two CBR modules to classify new cases and fetching obtained rules to combine the classification results. The whole process follows the medical protocol in this kind of decisions and the introduction of the rules guarantees more reliability in the integrated system because all rarities of the data are detected. Although the accuracy rates are in fact promising all these efforts are not enough in the most important study values for medical experts, the specificity and sensitivity rates. These results are crucial in medical research and need to tend to 100%.

Keeping in mind the goal of increasing the statistical results attached to false positives and negatives, we have intended to improve our work in a different manner. In this case instead of focusing our effort on improving the combination, issue accomplished with the D-module, we have worked with the case memory. Our wish has been the achievement of a better distance analysis between cases and, consequently, a more accurate classification. In this way, we have studied different pre-process techniques and, at last, we have completed our prior system (MEDIBE_II)[15] with a DML module. The decision of using DML is based on the characteristics of the domain and the improvements we want to obtain. All the details of this work and decisions will be explained in following sections.

The paper is organized as follows. Section 2 describes related works. Section 3 describes the new tool obtained by the combination of the different modules. In section 4, experimentation is presented and the performance is analysed. Finally, section 5 summarizes the conclusions and further work.

2 Related Work

The present work tries to classify new dermatological cases using the combination of different kinds of decisions. In the last years, the application of ensemble methods and collaborative systems have increased significantly. One of the most relevant lines is to improve clustering using ensembles [3]. There are also works to allow the classification using data of different complexity [2] and with different types of medical information [7]. In contrast to these approaches and the presence of several works focused on studying the melanoma domain from individual approaches such as in [6], there are not works to classify melanoma in a collaborative way that follow the medical roles and use independent CBR classifiers. We are working with two different points of view (confocal and dermatoscopic) from which we select the best classification from one of the systems depending on different criteria. These characteristics do not allow to talk about an Ensemble Learning system but the combination idea is similar enough to the one used in this kind of methods which combine the decisions from different systems to build a more reliable solution using the individual ones [12]. The combination of approaches

can be summarized [5] in: 1) Bagging, 2) Boosting and 3) Stacking. And the most common voting methods [10] are: 1) Plurality, 2) Contra-Plurality, 3) Borda-Count and 4) Plurality with Delete. All of them are based on the number of votes of a class (plurality). With all this information and attending to the medical necessities, the collaborative model is a good method for this kind of classification. In spite of it, the concrete characteristics of the domain [13] make necessary a method different from the classical ensemble ones. We are using different attributes of the same data in each system, then the independence of the data is guaranteed, in contrast to the standard Bagging. Analyzing the classification attributes, the vote method should be based on plurality but with some arranges requested by medical researchers, who weight more the information from Confocal Microscopy. Once we have experimented the protocol used by experts, we would like to improve it by using rules to do a better system combination.

In previous works, we have used clustering in order to discover new patterns on the medical domain [17] and we detected particular behaviours on the data. These characteristics guarantee a correct classification of the new patients when certain conditions in the attributes of the case were detected. Thus, bearing in mind that the main goals are the improvement of the classification and to minimize the false negatives situations, we created the D-Rule module. This module preprocesses the input data and creates a set of rules to help the whole classifier. Despite of using a similar idea of [9], we preprocess the data in a non-based interval way, where concrete values are detected and encapsulated in a rule. Moreover, our rules do not depend ones on each other and attributes are analysed independently.

In the field of Distance Metric Learning (DML) the methods could be divided into four families [11]. The first two are based on the supervision of the method. In this case, we have supervised and unsupervised DML. The last two families are based on a more concrete classification and are the DML methods based on Support Vector Machines and the ones that use Kernel methods. In our case and attending to the problem characteristics, we are working in the Supervised DML family [18]. With this methods we try to keep close all the data points within the same classes, while separating all the data points from different classes far apart. Even these techniques had been used and tested in clustering and retrieval, there are not literatures that applies it in a CBR environment.

3 Improving a Combination of CBR Systems in Dermatological Cancer

In this section, we want to describe the incremental improving of a basic classification system that tends to a complete Decision Support System for melanoma diagnosis. First of all, we describe the simple combination of CBR systems using the medical protocol. Then, we introduce the D-Rule module which through a preprocess algorithm obtain a better combination. And, in a third step we study the possibility of a projection of the original data in order to better classify the cases according to its similitude. At last, we show the result of the whole system.

3.1 Basic Combination of CBR Systems

Based on the medical protocol described in section one (shown at Fig. 1), we have developed a Collaborative Computer Aided System for melanoma diagnosis. In this way we have implemented an independent CBR system for each possible decision that uses the knowledge extracted from the Dermatoscopy and the Reflectance Confocal Microscopy image data. As a combination protocol, we follow in this case the same that medical experts implement in the daily diagnostic work [14].

In a more concrete way we can observe that the system binds the output of two Case-Based Reasoning (CBR) systems [1]. CBR performs the same resolution procedure that experts: solving new cases through the comparison of previous experiences. In a general way, the CBR life cycle can be summarized in the next four steps: 1) Retrieving the most similar cases from the case memory with the assistance of a similarity function, 2) Adapting the retrieved solutions to build a new solution for the new case, 3) Revising the proposed solution and 4) Retaining the useful knowledge generated in the solving process if it is necessary. An important point to highlight is that experts want to understand how decisions are made and this is one of the characteristics of CBR systems. Each one of the CBR systems has an independent case memory filled from the previously diagnosed injuries through the confocal and the dermatoscopy studies respectively. These two parts are completely independent and at the end of its work they put on its vote for the best classification according to their specific data. With this separate ballots, the system creates the final diagnosis (Solution) and, if proceeds, saves the new case in one of the case memories or in both.

The first option tested, as a decision process to perform a diagnosis, represents exactly the logical scheme used by the experts who mainly focus the interest on confocal diagnosis and, only if it is non conclusive they use the dermatological one. Therefore, the selection of the threshold values used to perform this decision and defined by medical experts are crucial to achieve a good performance. All this characteristics are described in the algorithm (Fig. 2).

Let c_{new} be the new input case
Let $best_{confocal}$ be the most similar case using the confocal CBR
Let $best_{dermatoscopical}$ be the most similar case using the dermatoscopical CBR
Let $distance(c_i, c_j)$ be the distance between two cases c_i and c_j performed by the normalized Euclidean distance
Let $threshold_{confocal}$ be the minimal value to accept two cases as similar from the confocal point of view
Let $threshold_{dermatoscopical}$ be the minimal value to accept two cases as similar from the dermatoscopical point of view
Let *class(c)* be the class of the case c
if *distance*(c_{new},$best_{confocal}$)< $threshold_{confocal}$ **then**
  return *class*($best_{confocal}$)
else
  if *distance*(c_{new},$best_{dermatoscopical}$) < $threshold_{dermatoscopical}$ **then**
    return *class*($best_{dermatoscopical}$)
  return *class*($best_{confocal}$)

Fig. 2. Algorithm to diagnose a new case using the confocal and dermatoscopical criteria with plain combination.

Let A be the set of all attributes of the medical domain
forall *attributes in A* **do**
  Let A_i be the attribute i of the set A
  Let V be the all possible values of attribute A_i
  forall *values in V* **do**
    if $\exists V_j$ || *class is unique for all cases in training set* **then**
      CreateRule (A,i,V,j,class)

Fig. 3. Algorithm to generate rules in the D-Rule system.

3.2 Rule-Based Combination of CBR Systems

This module is based in the analysis of the training data in order to obtain a set of rules. These rules summarize the data complexity in the case memory. The main goal is to represent the existing gaps in the data space with no information associated, in order to advise the classifier in this sense. On the same way, when the correct classification is guaranteed with a high reliability, an alert to the following classifiers is sent, if a set of characteristics in the input data is detected.

In our domain, it is quite usual that from a certain discrete value of an attribute, the variation of the final decision on the classification does not change. So, the intervals proposed in [9] have been eliminated and clear-cut zones affected always in the same way have been created. These zones are summarized in one or more rules. The algorithm followed to generate the rules (Fig. 3) shows that we use all possible values of an attribute to analyze the input data. One of the advantages found in the medical domain is that the attributes are well delimited, so it is difficult to find a new case with a different value of the predefined ones in the domain. The module output is a set of $if-then-else$ rules which quantity depends on the data complexity and varies with the different types of classification (Melanoma, Melanocytic or BCC).

To use this module we follow, mainly, the same scheme as in the basic approach using the best retrieved cases but adding the information provided by the rules. It allows the possibility of selecting automatically and better the best diagnosis. The logic process followed, as a combination, is the one described in Fig.4.

3.3 Application of Distance Metric Learning to the Case Memories

We would like to learn a metric that keep close all the data points from the same class and, at the same time, separate as far as possible the data points from different classes. This metric can be obtained

Let c_{new} be the new input case
Let $best_{confocal}$ be the most similar case using the confocal CBR
Let $best_{dermatoscopical}$ be the most similar case using the dermatoscopical CBR
Let $distance(c_i, c_j)$ be the distance between two cases c_i and c_j performed by the normalized Euclidean distance
Let $numRules_{confocal}$ be the number of rules carried out by $best_{confocal}$
Let $numRules_{dermatoscopical}$ be the number of rules carried out by $best_{dermatoscopical}$
Let *class(c)* be the class of the case c

```
if numRules_confocal > numRules_dermatoscopical then
    return class(best_confocal)
else
    if numRules_confocal < numRules_dermatoscopical then
        return class(best_dermatoscopical)
    else
        if distance(c_new,best_dermatoscopical) < distance(c_new,best_confocal) then
            return class(best_dermatoscopical)
        else
            return class(best_confocal)
```

Fig. 4. Algorithm to diagnose a new case using the confocal and dermatoscopical criteria and combining the results by rules.

from several methods but the one we have chosen is the one proposed by Xing [18]. In this approach we learn a global distance metric that minimizes the distance between pairs of data included in the equivalence constraints and data pairs from the inequivalence constraints are separated. To obtain this metric we apply the following process:

We have a set of data points C = {x_1,x_2,...,x_n} where n is the number of samples, each $x_i \in \mathbb{R}^m$ is a vector where m is the number of features. Then, the distance metric is the matrix A $\in R^{m\times m}$ and the distance between the data points expressed by

$$d_A^2(x, y) = \| x - y \|_A^2$$

So if we establish the equivalence constraint as

$$\mathrm{S} = \{x_i,x_j \mid x_i \text{ and } x_j \text{ belong to the same class}\}$$

and the inequivalence set as

$$\mathrm{D} = \{x_i,x_j \mid x_i \text{ and } x_j \text{ not belong to the same class}\}$$

we have to obtain A that

$$\min_{A\in R^{m\times m}} \sum_{(x_i,x_j)\in S} \| x_i - x_j \|_A^2$$

and

$$\mathrm{A} \preceq 0 \sum_{(x_i,x_j)\in D} \| x_i - x_j \|_A^2 \geq 1$$

Once we have obtained this matrix A, we could work with the metric to establish the distance between our cases. The calculations to obtain A could be done by different ways. In our case, we have focused our attention on Convex Programming [8] that provides an easy way.

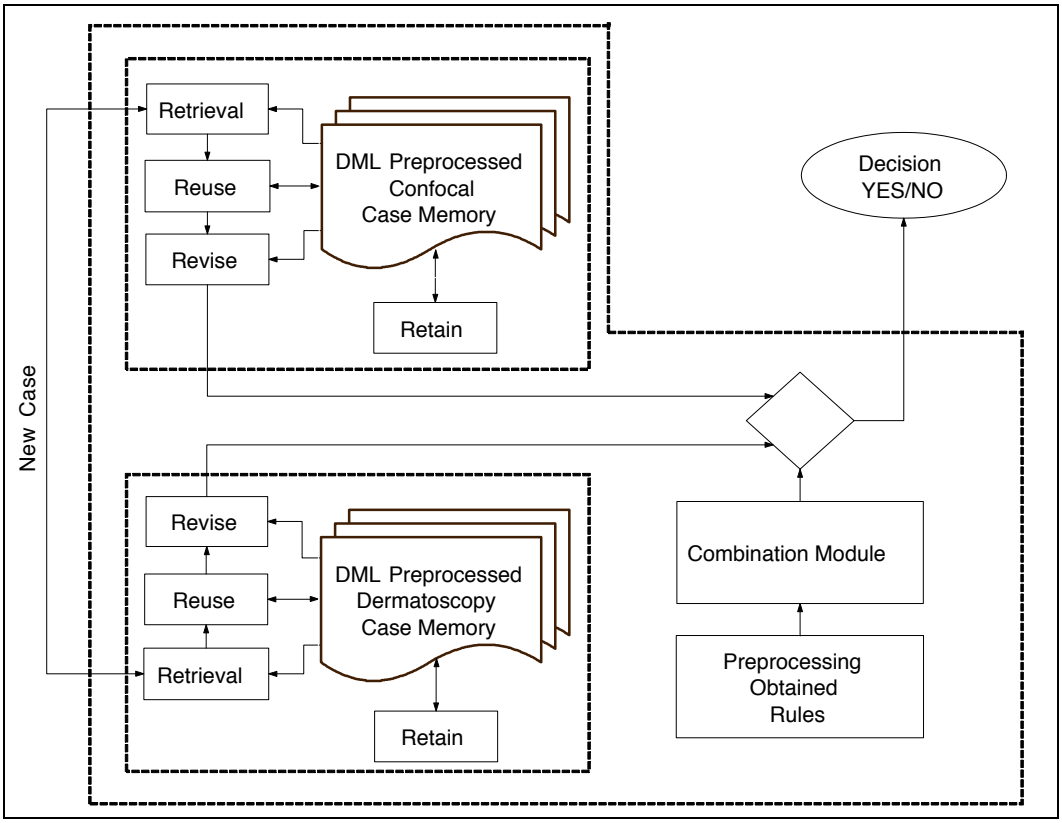

Fig. 5. Final Combined Decision Schema.

3.4 Rule-Based Combination of CBR Systems with Distance Metric Learning application to its Case Memories

The final approach used in this paper is to apply the DML techniques to the different case memories of each CBR system and then run these systems combining its results using the module based on preprocessing generated rules. This process is shown in figure 5. As we could see in this scheme we have two CBR modules one for confocal data and the other for the dermatoscopic one. For each one of these modules we have applied the Xing DML technique [18] in order to better distinguish among similar cases. The final classification is obtained through the most similar cases retrieved from each module and selected using the preprocessed rules.

4 Experimentation

This section describes the data extracted from images and analyzes the results of the experiments performed through sensitivity and specificity rates.

4.1 Testbed

One of the main difficulties in the classification of injuries in melanoma domain is the huge amount of information that new technologies are able to collect and the ignorance about how they are interrelated [13]. The most used techniques to gather information from tissue are the dermoscopic and the confocal analysis. Confocal microscope is the most precise and the one that medical experts consider as world class.

Nevertheless, a negative point is that the confocal analysis is a long and expensive test, so the number of available cases is limited. Due to this situation, the data set used in this work is composed only by 150 instances of suspicious lesions. All instances contain information related to confocal and dermatoscopic images and the histology corroborated diagnosis. Attending to the considerations of the

Table 1. Classification accuracy results.

	Melanoma	Melanocytic	BCC
Only Confocal Images	87%	90%	96%
Only Dermatoscopy Images	90%	98%	95%
Both Images with Plain Combination	89%	96%	95%
Both Images with Rules Combination	94%	99%	99%
Both Images with Distance Metric Learning	100%	100%	100%

Table 2. Sensitivity results.

	Melanoma	Melanocytic	BCC
Only Confocal Images	73%	94%	81%
Only Dermatoscopy Images	73%	99%	92%
Both Images with Plain Combination	70%	96%	92%
Both Images with Rules Combination	81%	100%	92%
Both Images with Distance Metric Learning	100%	100%	100%

medical experts that have created this set, it includes enough cases from each kind of illness to be representative of the domain. Then, in medical terms it is an appropriated case memory for this study. Detailing the instances, dermatological information has forty-one fields and confocal microscopy, due to its higher resolution, contributes with data from eighty-three different attributes.

4.2 Experimentation Framework

We have tested the classification accuracy of the platform proposed with a basic decision combination, with the use of rules obtained through the use of preprocessing algorithms and with the application of DML to the original data. In addition, we tested the accuracy of the two independent CBR systems (one for confocal data and another for dermatoscopy). This information is complemented with the analysis of the sensitivity and the specificity for the different cases. In the case of the plain combination platform, the medical consensus is to use 0.5 as confocal threshold and double of the distance between new case and best confocal case as dermatological one. All the CBR systems used in experimentation are configured with one-nearest neighbour algorithm with normalized Euclidean distance as retrieve function. This experiment framework has been tested applying a leave-one-out to the original data to obtain the average accuracy of those systems.

4.3 Results and Discussion

Analyzing these results, table 1 shows the accuracy rate to classify new injuries. This analysis is done in the three possible classes (Melanoma, Melanocytic and BCC) and using only confocal image, only dermoscopic image, both images with rules, without them, and with the DML module. The results obtained highlight two points: firstly, the techniques added to the basic combination of systems achieve better classification results than the plain combination or the independent systems. In a second term we could observe that the results using the whole system have the highest possible accuracy. On the other hand, tables 2 and 3 summarize the results of analyzing the statistics from the point of view of specificity and sensitivity, the most important values for medical experts. Analyzing the results from the more basic systems, we could see that it is more reliable to do a prognostic of real negative cases than the positive ones. This happens because data sets are unbalanced, what means that, they have different number of cases of each type because data sets represent a real situation: there are more

Table 3. Specificity results.

	Melanoma	Melanocytic	BCC
Only Confocal Images	92%	81%	99%
Only Dermatoscopy Images	96%	98%	95%
Both Images with Plain Combination	95%	96%	96%
Both Images with Rules Combination	98%	98%	100%
Both Images with Distance Metric Learning	100%	100%	100%

healthy people than ill people. Despite of it, the use of the combination of both types of images with the help of preprocessing obtained rules allows an important increase of sensitivity and specificity rates (in addition to improve the accuracy). It is important to highlight that, using the DML technique in order to better classify the new cases, we accomplish the desire of medical experts that is to avoid false negatives. This situation is shown with the 100% sensitivity and specificity for all the types of classification.

5 Conclusions and Further Work

As medical experts explain, early melanoma diagnosis is one of the main goals in dermatology. The diagnosis process with non-invasive techniques is complex because of the data typology. With our current work we have proposed a platform to aid experts through this diagnostic process in order to improve its classification results. The proposal combines information from techniques based on images through a combination algorithm based on experts' experiences. In addition, we use preprocessed rules in order to improve this combination and Distance Metric Learning to better analyze the distance between cases. After the result analysis of testing melanoma data sets, we can conclude that the incremental solutions to the problem allows an optimal classification even in specificity and sensitivity rates. These results also highlight that we have reached our ceiling with this data set. This situation makes us think about some important future challenges. The first of this is the obtaining of new data in this domain. Our previous works in data characterization of the domain shown that the new techniques used by medical experts generate high volumes of data. One of the main goals is to start using this data establishing relationships between it and extracting clear patters of its characteristics. So, we will need to work with those huge amounts of data in order to process them previous to start classification processes.

Acknowledgments

We would like to thank the Spanish Government for supporting the MID-CBR project under grant TIN2006-15140-C03, the Generalitat de Catalunya for the support under grants 2008SGR-00183 and 2009FI_B1 00092 and La Salle for supporting our research group. The work performed by S. Puig is partially supported by: FIS, under grant 0019/03 and 06/0265; Network of Excellence, 018702 GenoMel from CE.

References

1. A. Aamodt and E. Plaza. Case-based reasoning: foundational issues, methodological variations, and system approaches. *AI Communications*, 7(1):39–59, 1994.
2. R. Abdel-Aal. Abductive network committees for improved classification of medical data. In *Methods of Information in Medicine*, 2004.
3. R. Avogadri and G. Valentini. Fuzzy ensemble clustering for DNA microarray data analysis. In *Lecture Notes in Computer Science 4578*, 2007.
4. C. M. Balch. Prognostic factors analysis of 17,600 melanoma patients: Validation of the american joint committee on cancer melanoma staging system. In *Jounal Clinical Oncology*, pages 3622–3634. American Society Clinical Oncology, 2001.
5. E. Bauer and R. Kohavi. An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Machine Learning*, 36(1-2):105–139, 1999.
6. A. Fornells, E. Armengol, E. Golobardes, S. Puig, and J. Malvehy. Experiences using clustering and generalizations for knowledge discovery in melanomas domain. In *Advances in Data Mining. Medical Applications,E-Commerce, Marketing, and Theoretical Aspects*, volume 5077, pages 57–71. Springer, 2008.
7. A. Fornells, E. Golobardes, E. Bernadó, and J. M. Bonmatí. Decision support system for breast cancer diagnosis by a meta-learning approach based on grammar evolution. In *Eighth International Conference on EIS*, pages 222–229, 2006.
8. M. Grant, S. Boyd, and Y. Ye. Disciplined convex programming. In *Global Optimization: From Theory to Implementation, Nonconvex Optimization and Its Application Series*, pages 155–210. Springer, 2006.
9. T. Ho, M. Basu, and M. Law. Measures of complexity in classification problems. In *Data Complexity in Pattern Recognition, pp. 1-23, Springer*, 2006.
10. K. T. Leung and D. S. Parker. Empirical comparisons of various voting methods in bagging. In *Ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 595–600. ACM Press, 2003.
11. Y. Liu, Y. Yang, and J. Carbonell. Boosting to correct inductive bias in text classification. In *CIKM '02: Proceedings of the eleventh international conference on Information and knowledge management*, pages 348–355, New York, NY, USA, 2002. ACM Press.

12. P. Melville and R. J. Mooney. Diverse ensembles for active learning. In *ICML 04: Twenty-first international conference on Machine learning*, 2004.
13. R. Nicolas, E. Golobardes, A. Fornells, S. Puig, C. Carrera, and J. Malvehy. Identification of relevant knowledge for characterizing the melanoma domain. In *Advances in Soft Computing*, volume 49 2009, pages 55–59. Springer Berlin-Heidelberg, 2008.
14. R. Nicolas, E. Golobardes, A. Fornells, S. Segura, S. Puig, C. Carrera, J. Palou, and J. Malvehy. Using ensemble-based reasoning to help experts in melanoma diagnosis. In *Frontiers in AI and App., vol 184*, pages 178–185. IOS Press, 2008.
15. R. Nicolas, D. Vernet, E. Golobardes, A. Fornells, S. Puig, and J. Malvehy. Improving the combination of cbr systems with preprocessing rules in melanoma domain. In *Workshop Proceedings of the 8th International Conference on Case-Based Reasoning*, pages 225–234, 2009.
16. A. Scope and al. In vivo reflectance confocal microscopy imaging of melanocytic skin lesions. *J Am Acad Dermatol*, 57:644–658, 2007.
17. D. Vernet, R. Nicolas, E. Golobardes, A. Fornells, C. Garriga, S. Puig, and J. Malvehy. Pattern Discovery in Melanoma Domain Using Partitional Clustering. In *Frontiers in AI and App., 184, IOSPress*, pages 323–330, 2008.
18. E. P. Xing, A. Y. Ng, M. I. Jordan, and S. Russell. Distance metric learning, with application to clustering with side-information. In *Advances in Neural Information Processing Systems 15*, pages 505–512. MIT Press, 2002.
19. L. Yang. Distance metric learning: A comprehensive survey, 2006.

Using Ensemble-Based Reasoning to help experts in melanoma diagnosis. *Onzè Congrés Internacional de l'Associació Catalana d'Intel·ligència Artificial*, 2008

Using Ensemble-Based Reasoning to Help Experts in Melanoma Diagnosis

Ruben NICOLAS [a], Elisabet GOLOBARDES [a], Albert FORNELLS [a], Sonia SEGURA [b], Susana PUIG [c], Cristina CARRERA [c], Joseph PALOU [c], and Josep MALVEHY [c]

[a] *Grup de Recerca en Sistemes Intel.ligents*
Enginyeria i Arquitectura La Salle - Universitat Ramon Llull
Quatre Camins 2, 08022 Barcelona
{rnicolas,elisabet,afornells}@salle.url.edu

[b] *Dermatology Department of Hospital del Mar*
Passeig Marítim 25-29, 08003 Barcelona
ssegura@imas.imim.es

[c] *Melanoma Unit, Dermatology Department*
Hospital Clinic i Provincial de Barcelona, IDIBAPS
{spuig,ccarrera,jpalou,jmalvehy}@clinic.ub.es

Abstract. New habits in solar exposure cause an important increase of melanoma cancer during the last few years. However, recent studies demonstrate that early diagnosis drastically improves the treatment of this illness. This work presents a platform called MEDIBE for helping experts to diagnose melanoma. MEDIBE is an ensemble-based reasoning system which use two of the most important non-invasive image techniques: Reflectance Confocal Microscopy and Dermatoscopy. The combination of both image source improves the reliability of diagnosis.

Keywords. Artificial Intelligence in Medical Applications, Case-Based Reasoning, Ensemble Methods, Melanoma, Basal Cell Carcinoma, Reflectance Confocal Microscopy, Dermatoscopy.

Introduction

Death related to melanoma cancer has increased during the last few years due to the new solar habits. This growth convert it in a very important case study in our society, even more if we analyze that this kind of cancer affects people of any age, but especially young ones. According to the American Cancer Society, it is the skin cancer which causes most deaths because it is mortal in approximately twenty percent of cases [11,20]. Although an early diagnosis permits practically a secure regain, its identification is not trivial due to the different sizes, shapes and colors in which it can appear.

The Dermatology Department of *Hospital Clinic i Provincial de Barcelona* (HCPB) works with two of the most promising techniques of image analysis for melanoma diagnosis: Dermatoscopy and Reflectance Confocal Microscopy (RCM) [18]. The first is

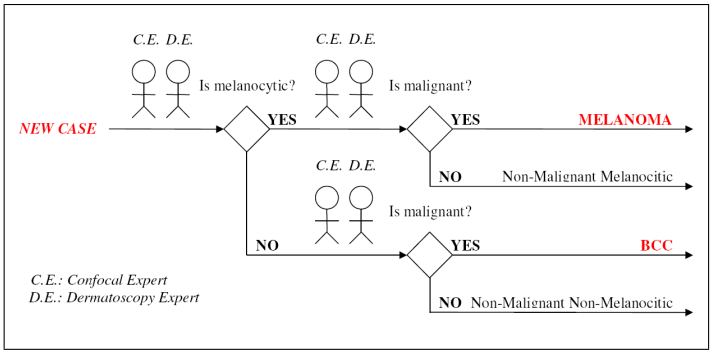

Figure 1. Medical protocol followed by experts from HCPB for melanoma diagnosis.

based in the microscopical image created with epiluminiscence (x10.30) and the second make the image with the reflectance of a coherent laser (x100) with a cell resolution [17]. Through the application of both analysis they perform a diagnostic process based on two step. First, they analyze if the new case is melanocytic or not. Next they assess if this case is malignant or not. Thus, the combination of both diagnosis allows experts to determine if the new case is Melanoma, Basal Cell Carcinoma (BCC) or a non-malignant tumor as figure 1 shows. Moreover, one benefit of this approach is that expert does not need to perform any invasive test to the patient.

The goal of this work is to develop a framework for helping experts to automate the diagnosis process under their supervision. The approach models the protocol described in figure 1 through several ensemble systems which contain the logical decision performed by Confocal and Dermoscopic experts. The ensemble is based in Case-Based Reasoning [1] because the approach uses past experiences to solve new cases and this is exactly the same procedure used by experts.

The paper is organized as follows. Section 1 describe some related work. Section 2 describe the medical application. Section 3 analyzes its performance. Section 4 ends with conclusions and further work.

1. Related Work

Ensemble methods combine the decisions from different systems to build a more reliable solution using the individual ones [12,13,15,19]. The combination of approaches can be summarized in: 1) Bagging, 2) Boosting, and 3) Stacking. Bagging [4] and Boosting [7,6] are based in the combination of the outputs using votes. In concrete Bagging replicates N systems of the same approach but using different data sources. In opposition Boosting follows the same idea but it define models in order to complement them. On the other hand Stacking [5] is based on heuristics that combine the outputs of several approaches. As voting methods the most common ones [14] are: 1) Plurality, 2) Contra-Plurality, 3) Borda-Count, and 4) Plurality with Delete. All these methods are based on the number of votes of a class (plurality) but with multiple types of addition of plurality and decision of better class.

Although there are works focused on studying the melanoma domain from individual approaches such as in [9], the application of ensemble methods has increased in last

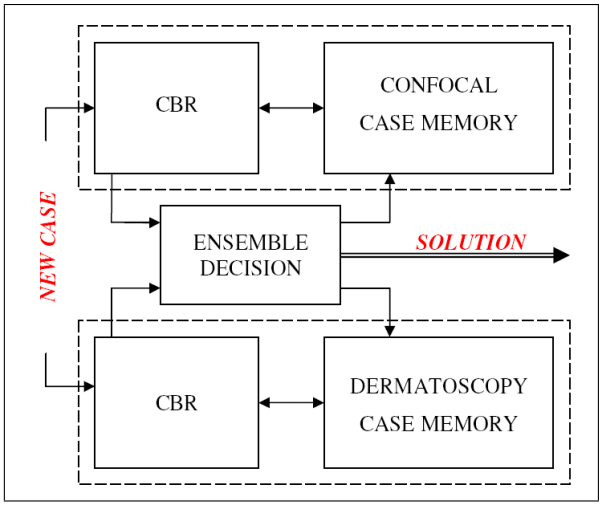

Figure 2. Ensemble Decision Schema.

years. One of this lines is to improve clustering using ensembles [3]. There are also works to allow the classification using data of different complexity [2] and with different types of medical information [8,10]. In contrast to these approaches we would like to classify in melanoma domain following the medical diagnosis protocol using different ensemble classifiers. Attending to this problem, the medical necessities and the existing data seems interesting to create an ensemble model with an expert for each kind of data. We note that we adapt the model using different attributes of the same data in each ensemble, then the independence of the data is guarantee, in contrast to the standard Bagging. Analysing that the classification attributes are boolean the vote method should be based on plurality but with some arranges requested by medical researchers, who weight more the information of an specific machine (Confocal Microscopy).

2. MEDIBE: A framework for melanoma diagnosis

MEDIBE (MElanoma DIagnosis Based on Ensembles) is a computer aided system for melanoma diagnosis based on the medical protocol described in the first section. For each one of the decision points, an ensemble system is used to answer the medical question using the knowledge extracted from the Dermatoscopy and the Reflectance Confocal Microscopy image data as Fig. 2 shows.

As we can observe, the ensemble system combines the output of two Case-Based Reasoning (CBR) systems [1]. This is because CBR performs the same resolution procedure than experts: solving new cases through the adaptation of previously solved cases. In a general way, the CBR life cycle can be summarized in the next for steps: 1) Retrieving the most similar cases from the case memory with the assistance of a similarity function; 2) Adapting the retrieved solutions to build a new solution for the new case; 3) Revising the proposed solution, and; 4) Retaining the useful knowledge generated in the solving process if it is necessary. Thus, the explanation capability is highly appreciated by the experts because they are able to understand how the decisions are done. Each one

Let c_{new} be the new input case
Let $best_{confocal}$ be the most similar case using the confocal CBR
Let $best_{dermatoscopical}$ be the most similar case using the dermatoscopical CBR
Let $distance(c_i, c_j)$ be the distance between two cases c_i and c_j performed by the normalized Euclidean distance
Let $threshold_{confocal}$ be the minimal value to accept two cases as similar from the confocal point of view
Let $threshold_{dermatoscopical}$ be the minimal value to accept two cases as similar from the dermatoscopical point of view
Let *class(c)* be the class of the case c
if *distance(*c_{new},$best_{confocal}$*)*<$threshold_{confocal}$ **then**
  return *class(*$best_{confocal}$*)*
else
  if *distance(*c_{new},$best_{dermatoscopical}$*)*<$threshold_{dermatoscopical}$ **then**
    return *class(*$best_{dermatoscopical}$*)*
  return *class(*$best_{confocal}$*)*

Figure 3. Algorithm to diagnose a new case using the confocal and dermatoscopical criteria.

of the CBR systems feed from two different case memories which stores all the previously diagnosed injuries through the confocal and the dermatoscopy studies respectively. These two parts are completely independent and at the end of its work they put on its vote for the best classification according to their specific data. With this separate ballots the system creates the final diagnosis (Solution) and, if proceed, save the new case in one of the case memories or both.

The decision process followed to perform a diagnosis is described in figure 3 and it represents the logical used by the experts. In spite of using a collaborative scheme where both diagnosis are combined, experts mainly focus on confocal diagnosis and, only if the diagnosis is non conclusive they use the dermatological diagnosis. Therefore, the selection of the threshold values used to decide if the relevance diagnosis is the confocal or the dermatological are crucial to achieve a good performance. Both values need to be defined by experts.

3. Experimentation

This section describes the data extracted from images and analyze the results of the experiments performed with MEDIBE through sensitivity and specificity rates.

3.1. Testbed

The classification of injuries in melanoma domain is not trivial. One of the main difficulties is the huge amount of information that new technologies are able to collect, and the ignorance about how they are related. One of the most used techniques to gather information from tissue is the dermoscopic analysis. Nevertheless, there are specific kinds of melanoma that it is not able to diagnose [16]. For this reason, experts want to evaluate if a new technique, called confocal analysis, is able to detect them or if the usage of both techniques can improve the individual analysis.

Table 1. Classification accuracy of MEDIBE using only confocal images, only dermoscopic images, and both images.

	Melanocytic	**Melanoma**	**BCC**
Confocal Image	89%	87%	96%
Dermatoscopy Image	88%	79%	90%
Confocal and Dermoscopic Image	95%	89%	97%

Table 2. T-test comparison between methods using 95% of confidence level

	Melanocytic	**Melanoma**	**BCC**
Ensemble - Confocal	↑	-	↑
Ensemble - Dermatoscopy	↑	↑	↑
Confocal - Dermatoscopy	-	↑	↑

Table 3. Sensitivity and specificity results from MEDIBE using combined information of confocal and dermoscopic images.

	Melanocytic	**Melanoma**	**BCC**
Specificity	91%	96%	100%
Sensitivity	96%	66%	88%

The dataset used in this work is composed by 150 instances of suspicious lesions from HCPB's patients. All instances contain information related to Confocal Image, Dermatoscopy Image and Diagnosis (corroborated with the histology). Attending the considerations of the medical experts that have created this set, it includes enough cases from each kind of illness (or healthiness) to be representative of the domain. Then, in medical terms it is an appropriated case memory for this study. Detailing the instances, dermatological and confocal image data are collected from two different microscopes with different kinds of precision. The first one, known as Dermatoscopy, have forty-one fields with knowledge of Symmetry, Color, Reticle, Globules, and other concrete aspects [16]. The second one, called Reflectance Confocal Microscopy (RCM), have data from eighty-three different attributes with information about Vessels, Nucleated and Non-Nucleated Cells, Dermal Papilla, Pagetoid Infiltrations, Basal Cell, and other [17]. RCM contributes with more attributes due to its higher resolution than Dermatoscopy.

Figure 4 contains images of four different types of injuries: A) Melanoma in situ, B) Melanoma, C) Nevus (melanocytic non-melanoma), and D) Basal Cell Carcinoma. For each one of types, different level of details are available: 1) Clinical image, 2) Dermatoscope image, 3) In vivo Confocal image, and 4) Histo-pathological study with hematoxylin eosine stain. It is important to highlight that the different types of image information could give very different aspects even in same final diagnostic cases. This situation is clear comparing A and B part where we have two melanomas but in first case A the lesion studied is clinically and dermoscopically banal (non-suspicious of malignancy) but confocal is characteristic of melanoma as is demonstrated in the histology. In case B the diagnosis of melanoma is clear in all images. Part C and D permits to see that in non-melanoma melanocytic cases and in BCC ones clinical images are similar but under dermatoscopy and confocal microscopy have specific criteria to reach the correct diagnosis, nevus (benign melanocytic lesion) and BCC (malignant non-melanocytic).

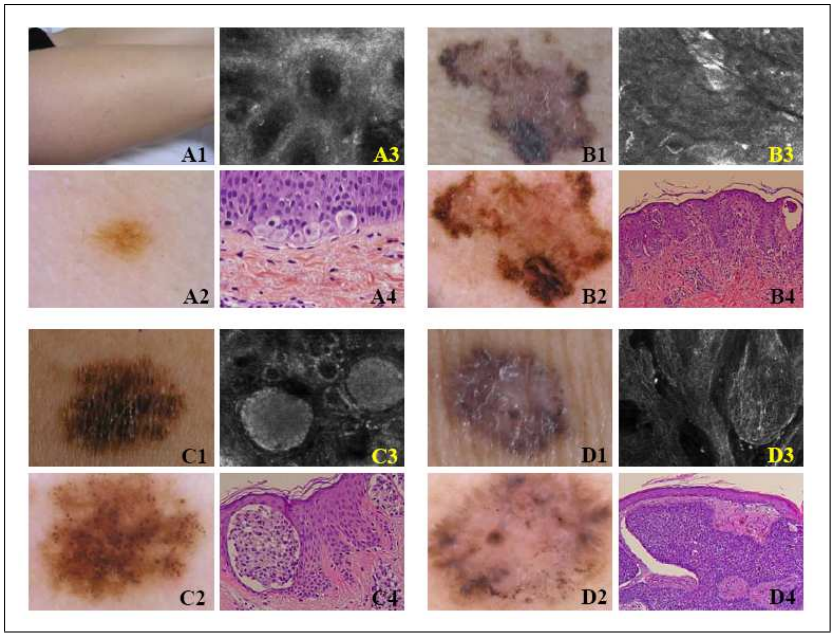

Figure 4. Images from cases included in the study.

3.2. Experimentation, Results and Discussion

We have test the classification accuracy of MEDIBE and the one of independent CBR systems (one for confocal data and another for dermatoscopy). Both experimentations uses a CBR configured with one-nearest neighbor algorithm with normalized Euclidean distance as retrieve function. We have test different confocal and dermatoscopical thresholds according to medical experts criteria. Finally we have reach a consensus to use 0.5 as confocal threshold and double of the distance between new case and best confocal case as dermatological one. With this experiment framework we have applied a 10-fold stratified cross validation to the original data to obtain the average accuracy of those systems.

Analyzing these results, table 1 shows the accuracy rate to classify new injuries in the three possible classes (Melanocytic, Melanoma and BCC) using only confocal image, only dermoscopic image and both images. To evaluate the representativity of accuracy differences table 2 shows the results obtained from the application of t-test (at 95% of confidence level) to the accuracy results of different data configurations. We represent with the symbol (↑) if the first method is significantly better than the second one, and with the symbol (-) if its significance is not representative. Results of accuracy and significance, highlight two points: First, confocal analysis diagnoses better than dermatoscopy (with significant difference in all cases except in melanocytic differentiation where are equivalent). Second, the combination of both data improve the results (despite of the case of Melanoma classification where the use of only confocal information is equivalent to the combined one).

On the other hand, table 3 summarizes the results of analysing the statistics from the point of view of specificity and sensitivity. They show that is more reliable to do a

prognostic of real negative cases than the positive ones. This fact is because datasets are unbalanced, that is, they have different number of cases of each type because datasets represent a real situation: there are more healthy than sick people.

Detailing negative cases, we could say that in melanocytic distinguish the percentage of bad classification is nine percent, in melanoma four percent, and at last in BCC the system does not fail in any case. Notwithstanding positive cases classification give worse results, the sensitivity of classification are better than the obtained without the application of the proposed method. As *e.g.*, doing the analysis of the worst sensitivity result of the system (66% obtained in melanoma classification) it is better than the 57% that have medical researchers in difficult melanoma classification cases nowadays.

4. Conclusions and Further Work

Melanoma diagnosis using non-invasive techniques is nowadays one of the most important goals in dermatology, due to the necessity of getting an early diagnosis. The diagnostic process is complex because of the highly variability of shapes, sizes, and colors.

We propose a platform called MEDIBE for automatizing medical protocol followed by HCPB's experts to diagnose a melanoma. The application combines information from two of the most promising techniques based on images through an ensemble algorithm founded on experts' experiences. After the analysis of results from testing melanoma dataset, we can conclude that the combination of both images improves the individual results applying the medical protocol.

The further work is focused on two lines. First, create a new specific voting method not conditioned by medical criteria. This independence with experts could permit to find aspects not included in formal algorithms but used unconsciously by them, as clinical details. Second, it would be useful to test a different ensemble method idea in diagnostic. It is to create a system that permits to prognostic if a lesion is malignant or not with independence to its concrete type. Then, the different ensembles could be trained to vote according to specific problem (melanocytic, melanoma, BCC, and other) and not in reference to one type of image. Finally, this ensemble has to vote if the new case needs treatment or not.

Acknowledgements

We thank the Spanish Government for the support in MID-CBR project under grant TIN2006-15140-C03 and the Generalitat de Catalunya for the support under grants 2005SGR-302 and 2008FI_B 00499. We thank Enginyeria i Arquitectura La Salle of Ramon Llull University for the support to our research group. The work performed by S. Puig and J. Malvehy is partially supported by: Fondo de Investigaciones Sanitarias (FIS), grant 0019/03 and 06/0265; Network of Excellence, 018702 GenoMel from the CE.

References

[1] A. Aamodt and E. Plaza, Case-Based Reasoning: Foundations Issues, Methodological Variations, and System Approaches, *AI Communications* **7** (1994), 39–59.

[2] R.E. Abdel-Aal, Abductive network committees for improved classification of medical data, *Methods of Information in Medicine* **43** (2004), 192–201.

[3] R. Avogadri and G. Valentini, Fuzzy ensemble clustering for DNA microarray data analysis, *LNCS 4578 LNAI* (2007), 537–543.

[4] E. Bauer and R. Kohavi, An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants, *Machine Learning Journal* **36** (1999), 105–139.

[5] B. Clarke, Comparing Bayes model averaging and stacking when model approximation error cannot be ignored, *J. Mach. Learn. Res.* **4** (2003), 683–712.

[6] Y.S. Dong and K.S. Han, Boosting SVM classifiers by ensemble,*WWW '05: Special interest tracks and posters of the 14th International Conference on World Wide Web* (2005), 1072–1073.

[7] G. Eibl and K.P. Pfeiffer, Multiclass Boosting for Weak Classifiers, *J. Mach. Learn. Res.* **6** (2005), 189–210.

[8] A. Fornells, E. Golobardes, E. Bernadó and J. Martí, Decision Support System for Breast Cancer Diagnosis by a Meta-Learning Approach based on Grammar Evolution. *9th International Conference on Enterprise Information Systems* (2006), 222–227.

[9] A. Fornells, E. Armengol, E. Golobardes, S. Puig and J. Malvehy, Experiences Using Clustering and Generalizations for Knowledge Discovery in Melanomas Domain. *7th Industrial Conference on Data Mining* **5077** (2008), 57–71.

[10] Y. Gao et al., LCSE: Learning classifier system ensemble for incremental medical instances, *LNCS 4399 LNAI* (2007), 93–103.

[11] R.M. Gutierrez and N. Cortes, Confronting melanoma in the 21st century, *Med Cutan Iber Lat Am* **35** (2007), 3–13.

[12] D.A. Hull, J.O. Pedersen and H. Schütze, Method combination for document filtering, *SIGIR '96: Proceedings of the 19th Annual International ACM SIGIR Conference on Research and development in information retrieval* (1996), 279–287.

[13] N. Indurkhya and S.M. Weiss, Solving regression problems with rule-based ensemble classifiers, *KDD '01: Proceedings of the seventh ACM SIGKDD International Conference on Knowledge discovery and data mining* (2001), 287–292.

[14] K.T. Leung and D.S. Parker, Empirical comparisons of various voting methods in bagging, *KDD '03: Proceedings of the ninth ACM SIGKDD International Conference on Knowledge discovery and data mining* (2003), 595–600.

[15] P. Melville and R.J. Mooney, Diverse ensembles for active learning, *ICML '04: Twenty-first international conference on Machine learning* (2004).

[16] S. Puig et al., Melanomas that failed dermoscopic detection: a combined clinicodermoscopic approach for not missing melanoma, *Dermatol Surg* **33** (2007), 1262–1273.

[17] A. Scope et al., In vivo reflectance confocal microscopy imaging of melanocytic skin lesions: Consensus terminology glossary and illustrative images, *J Am Acad Dermatol* **57** (2007), 644–658.

[18] S. Segura et al., Dendritic cells in pigmented Basal cell carcinoma: a relevant finding by reflectance-mode confocal microscopy, *Arch Dermatol* **143** (2007), 883–889.

[19] W.N Street and Y.S. Kim, A streaming ensemble algorithm (SEA) for large-scale classification, *KDD '01: Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining* (2001), 377–382.

[20] M.A. Tucker and A.M. Goldstein, Melanoma etiology: where are we?, *Oncogene* **22** (2003), 3042–3052.

Identification of relevant knowledge for characterizing the melanoma domain. *Onzè Congrés Internacional de l'Associació Catalana d'Intel·ligència Artificial*, 2008

Pattern Discovery in Melanoma Domain Using Partitional Clustering

David VERNET [a], Ruben NICOLAS [a], Elisabet GOLOBARDES [a], Albert FORNELLS [a], Carles GARRIGA [a], Susana PUIG [b] and Josep MALVEHY [b]

[a] *Grup de Recerca en Sistemes Intel·ligents*
Enginyeria i Arquitectura La Salle, Universitat Ramon Llull
Quatre Camins 2, 08022 Barcelona (Spain)
{dave,rnicolas,elisabet,afornells,cgarriga}@salle.url.edu

[b] *Melanoma Unit, Dermatology Department*
IDIBAPS, U726 CIBERER, ISCIII
Hospital Clinic i Provincial de Barcelona (Spain)
{spuig,jmalvehy}@clinic.ub.es

Abstract. Nowadays melanoma is one of the most important cancers to study due to its social impact. This dermatologic cancer has increased its frequency and mortality during last years. In particular, mortality is around twenty percent in non early detected ones. For this reason, the aim of medical researchers is to improve the early diagnosis through a best melanoma characterization using pattern matching. This article presents a new way to create real melanoma patterns in order to improve the future treatment of the patients. The approach is a pattern discovery system based on the K-Means clustering method and validated by means of a Case-Based Classifier System.

Keywords. Melanomas, Pattern Discovery, Clustering, Artificial Intelligence in Medicine, Computer Aided Systems, Case-Based Reasoning.

Introduction

New social habits in solar exposure make much more important the melanoma early diagnosis due to its high ratio of people affected by this illness . Analyzing data from the American Cancer Society and from specialists in skin cancer we could observe that this type of disease is increasing its rates of appearance. In addition, this kind of skin cancer is which accumulate highest percentage of death in front of more common dermatological cancers, approximately a twenty percent of non early prognosticated cases [9]. Recently diagnosis of melanoma is based on the ABCD rule [7] which considers the following features commonly observed in this kind of tumour: (1) a diameter larger than 5 mm (2) colour variegation (3) asymmetry and (4) border irregularity. Although most of melanomas are correctly diagnosed following this rule, a variable proportion of cases does not comply with these criteria. Therefore, medical researchers focus on identifying reliable melanoma patterns to improve diagnosis because the actual patterns do not offer a good one.

The goal of this paper is to identify a set of reliable patterns from a dataset built by a group of melanomas seniors researchers. The extraction and validation process is performed by ULIC (Unsupervised Learning in CBR) system. The platform explores data from a data intensive approach based on partitional clustering technique.

Clustering algorithms can be classified in two basic types: hierarchical and partitional clustering. Hierarchical clustering proceeds successively by either merging smaller clusters into larger ones, or by splitting large clusters. The different algorithms based on this type of clustering differ in the way which two small clusters are merged or which large cluster is split. On the other hand, partitional clustering attempts to directly decompose the data set into a set of disjoint clusters. The choice of clustering algorithm depends on the type of data available and on the particular purpose and application [10]. In our case, we want to identify disjoined regions in the domain associated to a certain pattern. For this reason partitional algorithms are the most suitable for this purpose.

On the hand, the validation process is assessed through a Case-Based Reasoning (CBR) [1] approach. The main benefit of this approach is its capability for explaining why the classification has been done. Therefore, this issue is crucial helping experts to understand the results.

This paper is organized as follows. In section 1 the related work is presented. Section 2 describes the framework. Section 3 summarizes the experimentation and results. Section 4 exposes the conclusions and further work.

1. Related Work

Clustering techniques are a smart way to extract relationships from huge data amounts. Consequently, this useful property has been widely used in medical domains [6]. The types of work mainly depend on the data topology and the usage of extracted relations from analysis. As we have commented in previous section the Partitional Clustering is the most suitable approach for our purpose.

Two points of view can be taken in the Partitional clustering. The first one is the hard clustering, where each case is assigned to only one cluster. Second one, is the fuzzy clustering where each case is assigned a degree of membership of between 0 and 1 to each cluster. One of the main partitional hard clustering technique is the K-means algorithm [13]. There are special variations to improve some aspects of the algorithm. One of these is the K-medoids algorithm or PAM (Partition Around Medoids) [12].

There are melanoma studies focused in the identification of relationships between malignant melanoma and familiar or hereditary tumours (i.e. breast cancer, ovarian cancer, colon cancer, pancreatic cancer) such as in [15]. On the other hand, others works analyse thousands of genes with the aim of extracting the 'guilty' genes [5,4] related to the cancer. Anyway, both approaches help experts to be aware and detect melanoma formation in early stages.

Some works, like [11], have studied techniques available for the validation of clustering results. In particular, in [2] the standard measure of interset distance (the minimum distance between points in a pair of sets) was computed. It was the least reliable measure and experimental results also suggest that intercluster separation plays a more important role in cluster validation than cluster diameter. Dunn's indexes provided good validation results.

We remark the importance of CBR in Computer Aided Systems for medical investigation, as has been studied in psychiatry [3], renal failure cases [14], breast cancer [8] and so on. But nowadays it seems interesting to apply it to melanomas cancer attending to its increasing importance.

2. Discovering melanoma patterns with ULIC platform

One of the applications of ULIC (Unsupervised Learning In Case-Based Classifier Systems) platform (author cited) is to discover patterns and evaluate them. It contains a set of unsupervised methods which allow to explore non labelled data in order to create groups of similar examples represented by a prototype. Each prototype is what system considers as one possible class. In this paper, the aim is to extract patterns with usefulness for medical researchers based on the clusterization of melanomas information. Next, a second phase using CBR is applied in order to validate these patterns.

Medical researchers want to know what is the optimal number of melanoma patterns and characterize each one of them. For this reason, we have tackled the patterns discovery process with the K-means algorithm [13] because it allows us to perform an incremental search using different number of clusters, even other approach could be tested. Next, with the aim of checking the quantitative goodness of these new classifications, we use CBR system to assess the performance of the patterns discovered.

Next sections introduce CBR, k-means algorithm, and how they are integrated.

2.1. Case-Based Reasoning

CBR is a technique that solves new cases using others previously solved. In order to get this objective four phases are applied [1]. 1) Retrieve the most similar cases from the case memory with the assistance of a similarity function. 2) Try to reuse the solutions from the retrieved cases with the aim to solve the present case. 3) Make a revision of the solution. 4) Retain the useful information of the solved case, if it is necessary.

All steps turns around the Case Memory, which contains the experience of system in terms of cases. A case is a description of a problem.

2.2. K-means clustering

K-means is an unsupervised clustering method [13] that permits to do a partition of the domain in K clusters. The algorithm can be described as follows:

1: Choose an initial partition of the cases into k clusters. This is random assignment to k clusters.
2: Compute the distance from every case to the mean of each cluster and assign the cases to their nearest clusters.
3: Recompute the cluster means following any change of cluster membership at step 2.
4: Repeat steps 2 and 3 until no further changes of cluster member-ship occur in a complete iteration. The procedure has now converged to a stable k-partition.

At the end of the process, the mean value of the cases assigned to the cluster defines the centroid of the cluster, which represents the general properties of cases.

The distance function used in k-means is based on the Euclidean distance. For the discrete attributes, the most repeated value is used as mean value and the Heterogenous Value Difference Metric (HVDM) [16].

In order to use the Euclidean distance we get the assumption that the attributes are statistically independent and they vary in an equal way in clusters with a hiperspheric shape. More over, all clusters contain at least one case, so no empty clusters are obtained after applying the algorithm.

2.3. Integration of K-means and CBR in ULIC

ULIC is a system capable of discovering patterns in a specific domain. The platform integrates a system for exploring the different groupings with different number of clusters (K-means) and another one for evaluating the possible solutions we have obtained (CBR) (see Fig. 1).

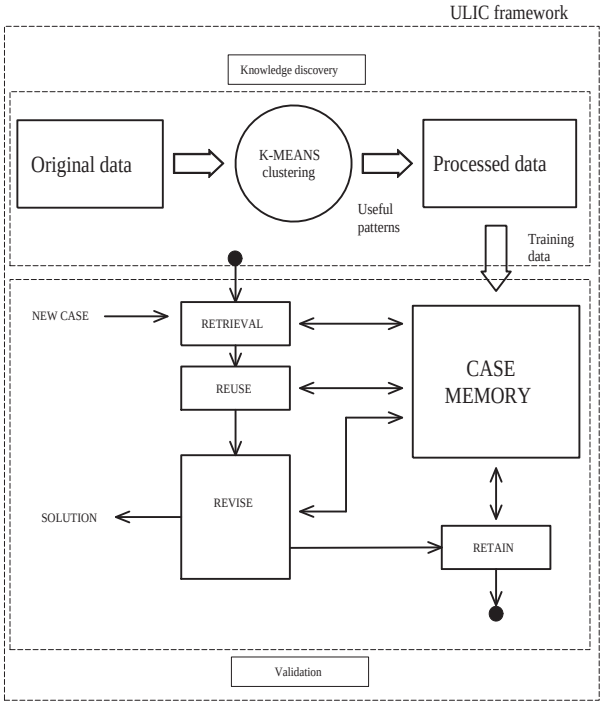

Figure 1. ULIC framework.

Let us have a look to each system module. First of all, we find the module where the data is processed. At this point, K-means clustering is applied to input data in order to discover new patterns in data structure. Using the clusters obtained by this module, the training cases are redefined and their classes are changed to the new ones.

The second module consists of a traditional CBR classifier that uses the 4-R cycle. According to the cluster in which they are the new data organization is introduced in the

Case Memory and the new training set is generated. CBR is used so as to validate the new clusters obtained in the case memory.

It is important to notice that it is not necessary to apply the CBR cycle when the medical staff need information about the patterns. Using the same technique, they could extract this information from the Memory Case using the previously classified cases.

An aspect to comment is that the labels of each cluster are unknown and it is not possible to generate a training data set and a test data set with the original data without avoiding the loss of information. Clustering is applied to whole data.

3. Experiments: Results and Discussion

As we have explained in previous sections, the aim of this work is to discover new patterns related to the melanoma domain to help medical staff in the early diagnosis. The next sections describe the dataset defined by medical experts and the melanoma patterns defined.

3.1. Data Description

The medical data used in our experimentation is extracted using real cases from the Melanoma Unit in *Hospital Clinic i Provincial de Barcelona*. The kind of data we have used is from two types: Dermatological, which represents the information gathered by dermatologists (that represents very specific knowledge) and Histological, which is the data collected by nurses (include more generic information). On the other hand, we could see that dataset is very varied and includes numerical and categorical values. We would like to remark that the attributes with categorical values, in spite of being integer values, can not be ordered for its value because it works as a label. The concrete domain and data type for each attribute are shown at Table 1. Although dataset is small (70 instances), it is considered by experts as an enough representation.

The next section focus on defining new classification for each sample.

Table 1. Attributes Domain and Types.

Attribute	Domain	Type
sex	Male, Female	histological
age	numeric	histological
max diam	numeric	histological
site	trunk,lower_extr,upper_extr,back,leg, arm,forearm,sovraclav,neck,shoulder	histological
pigment network	0,1,2	dermatological
dots and globules	0,1,2	dermatological
streaks	0,1,2	dermatological
regression structures	0,1,2,3,4	dermatological
BW veil	0,1	dermatological
bloches	0,1,2	dermatological
vessels	0,1,2,3,4	dermatological
millia like cyst	0,1	dermatological

3.2. Discovering New Patterns

With the aim of improving the possibility of getting a right diagnosis, we have used an incrementally exploration of K in the K-means algorithm. This incremental execution involves the analysis of classification between two and twenty different classes (clusters) and ten seeds for each one. The experts considered that 20 classes were sufficient for a so low number of instances.

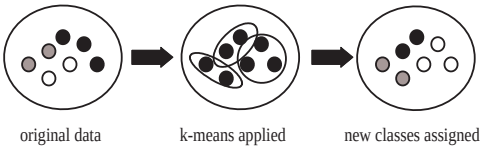

Figure 2. New dataset creation process. Cases are relabelled according to the cluster in which they belong

After having these results, for each seed from each number of classes, we have created a new dataset with the same instances as the initial one but with a new class attribute, attending to the cluster it has been assigned in(Fig. 2) previously. By the use of this new set, we have studied the percentage of the error rate of the CBR classification. The validation of the new classes is based on the execution of a CBR cycle setting Nearest Neighbor as retrieval phase and leave-one-out (due to the lack of original cases).

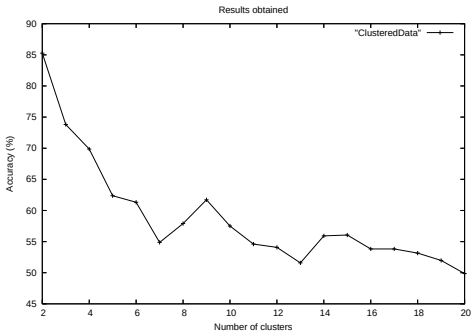

Figure 3. Results obtained using the artificial clusters. Ten different seeds are used for each number of clusters. Mean accuracy is represented for each *k*.

Fig. 3 summarizes the results as the average accuracy rate of the different seeds of every classification. The analysis of the results shows that the classification pattern, obtained with the K-means part of ULIC, permits an accuracy of at least fifty percent. The superior limit of correctly classification is around eighty-five percent. Both superior and inferior limits correspond to the smaller number of cases (2 clusters) and the biggest one (20 clusters), respectively.

The decreasing tendency has an exception for nine clusters, when an increase of approximately ten percent of accuracy is obtained.

In summary, we could say that focusing the attention on a quantitative analysis, we could remark four important ideas: 1) The classification results are better with not

many clusters. 2) We have a local maximum with nine clusters. 3) The results remark a decreasing tendency. 4) A recommendation of new classification criteria is to use nine classes.

Analyzing the patterns obtained by a 9-Means classification we can observe the following information (just informative):

- Cluster 0: The attributes $bloches$ and $millia\_like\_cyst$ are always 0. The site is always different of $shoulder$, $neck$, arm or $upper - extr$.
- Cluster 1:The attribute Max_diam is never large or huge. The attributes $pigment\_network$, $streaks$, BW_veil and $millia\_like\_cyst$ are always 0. $Dots\_and\_globules$ is always 2.
- Cluster 3: $Vessels$ is always different of 2, 3 and 4. The $site$ is different of sovraclav and shoulder.
- Cluster 4: The sex is always male. None are young. BW_veil, $bloches$, $vessels$ and $millia\_like\_cyst$ are always 0. $Pigment\_network$ is always 2.
- Cluster 5: BW_veil and $millia\_like\_cist$ are always 0. Attributes $Bloches$ and $Pigment\_network$ are always different of 0 and 2 respectively.
- Cluster 6: $Regression\_structures$ and BW_veil are always 0. $Streaks$ is always different of 1. $Dots\_and\_globules$ is always different of 2.
- Cluster 7: All are male. $Dots\_and\_globules$, BW_veil, $bloches$, $vessels$ and $millia\_licke\_cyst$ are always 0. None are huge or large (Max_diam).
- Cluster 8: All are male. $Pigment\_network$ is always 2. BW_veil, $vessels$, $millia\_like\_cyst$ are always 0. Max_diam is always different of small and huge.

Attending to the evaluation performed by the medical staff, which is a 8-classes classification, we propose to analyze now the possibility of classifying with nine classes instead of the eight used until now.

A preliminary study performed by medical staff of these clusters showed interesting groups of data. This study has to be complemented subsequently by medical researchers in a medical way, adding meaning to the patterns we have extracted.

4. Conclusions and Further Work

Due to new habits in solar exposure, melanoma is increasing its appearance. Early diagnosis is crucial in order to assault the recovery with successful possibilities. In this way it is important to tackle the search of accurate melanoma patterns. Achieving these patterns could permit non expert doctors to easy recognize suspicious injuries. This work uses ULIC to extract these patterns analyzing real medical data. ULIC is a platform for discovering data patterns through clustering techniques. Moreover, ULIC is capable to validate the extracted patterns using a CBR system. The main benefit of CBR is its capability for explaining the results to the experts.

The results of the experimentation provide experts an encouraging point of view in the discovering of melanoma patterns due to the promising results achieved.

The further work is focus on analyzing the medical utility of the new patterns. Medical staff should now analyze the possibility of classifying with nine classes instead of the eight used till now. At this point, we should corroborate the patterns with the comments of senior medical researchers.

Acknowledgements

We would like to thank the Spanish Government for the support in MID-CBR project under grant TIN2006-15140-C03 and the Generalitat de Catalunya for the support under grants 2005SGR-302 and 2007FI-A 01328. Also, we would like to thank Enginyeria i Arquitectura La Salle of Ramon Llull University for the support to our research group. We also would like to thank to the clinicians involved in the confection of the dataset: Dr Paolo Carli (dermatologist), Dr Vinzenzo di Giorgi (dermatologist), Dr Daniela Massi (dermatopathologist) from the University of Firenze; Dr Josep Malvehy (dermatologist), Dr Susana Puig (dermatologist) and Dr Josep Palou (dermatopathologist) from Melanoma Unit in Hospital Clinic i Provincial de Barcelona. Part of the work performed by S. Puig and J. Malvehy is partially supported by: Fondo de Investigaciones Sanitarias (FIS), grant 0019/03 and 06/0265; Network of Excellence, 018702 GenoMel from the CE.

References

[1] A. Aamodt and E. Plaza. Case-Based Reasoning: Foundations Issues, Methodological Variations, and System Approaches. In *AI Communications*, volume 7, pages 39–59, 1994.

[2] J. C. Bezdek and N. R. Pal. Some new indexes of cluster validity. 28(3):301–315, June 1998.

[3] I. Bichindaritz. A case-based assistant for clinical-psychiatry expertise. *Journal Of The American Medical Informatics Association*, pages 673–677, 1994.

[4] H. Dang, T. Segaran T. Le, and J. Levy. Integrating database information in microarray expression analyses: Application to melanoma cell lines profiled in the nci60 data set. *Journal of Biomolecular Techniques*, 13:199–204, 2002.

[5] M.B. Eisen, P.T. Spellman, P.O. Brown, and D. Botstein. Cluster analysis and display of genome-wide expression patterns. *National Academy Scientific*, 25(95):14863Ű14868, 1998.

[6] A. Fornells, E. Armengol, E. Golobardes, S. Puig, and J. Malvehy. Experiences using clustering and generalizations for knowledge discovery in melanomas domain. In *7th Industrial Conference on Data Mining*, volume 5077 of *LNCS*, pages 57–71. Springer-Verlag, 2008.

[7] R. J. Friedman, D. S. Rigel, and A. W. Kopf. Early detection of malignant melanoma: The role of physician examination and self-examination of the skin. *Ca-A Cancer J Clinicians*, 35:130–151, 1985.

[8] E. Golobardes, X. Llora, M. Salamo, and J. Martí. Computer Aided Diagnosis with Case-Based Reasoning and Genetic Algorithms. In *Journal of Knowledge-Based Systems*, volume 15, pages 45–52, 2002.

[9] R.M. Gutiérrez and N. Cortés. Confronting melanoma in the 21st century. In *Med Cutan Iber Lat Am*, pages 35(1):3–13, 2007.

[10] J. Han and M. Kamber. Data mining: Concepts and techniques, 2006.

[11] J. Handl, J. Knowles, and D.B. Kell. Computational cluster validation in post-genomic data analysis. *Bioinformatics*, 21(15):3201–3212, 2005.

[12] L. Kaufman and P.J. Rousseeuw. Clustering by means of medoids. In *Statistical Data Analysis Based on the L1-Norm and Related Methods*, pages 405–416, North-Holland, 1987. Y. Dodge.

[13] J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. Vol. 1.*, pages 281–297, 1967.

[14] S. Montani, L. Portinale, G. Leonardi, and R. Bellazzi. Case-based retrieval to support the treatment of end stage renal failure patients. *Artificial Intelligence in Medicine*, 3:31–42, 2006.

[15] P. Stefano, G. Fabbrocini, M. Scalvenzi, B. Emanuela, and M. Pensabene. Malignant melanoma clustering with some familiar/hereditary tumors. *Annals of Oncology*, 3:100, 2002.

[16] D. Randall Wilson and Tony R. Martinez. Improved heterogeneous distance functions. *Journal of Artificial Intelligence Research (JAIR)*, 1:1–34, 1997.

Pattern discovery in melanoma domain using partitional clustering. *2nd. International Workshop on Practical Applications of Computational Biology and Bioinformatics*, 2008

Identification of Relevant Knowledge for Characterizing the Melanoma Domain

Ruben Nicolas[1], Elisabet Golobardes[1], Albert Fornells[1], Susana Puig[2], Cristina Carrera[2], Josep Malvehy[2]

1 Grup de Recerca en Sistemes Intel·ligents, Enginyeria i Arquitectura La Salle, Universitat Ramon Llull, Quatre Camins 2, 08022, Barcelona, Spain.

2 Melanoma Unit, Dermatology Department, Hospital Clinic i Provincial de Barcelona, IDIBAPS, Barcelona, Spain.

{rnicolas,elisabet,afornells}@salle.url.edu, {spuig,ccarrera,jmalvehy}@clinic.ub.es

Abstract. Melanoma is one of the most important cancers to study in our current social context. This kind of cancer has increased its frequency in the last few years and its mortality is around twenty percent if it is not early treated. In order to improve the early diagnosis, the problem characterization using Machine Learning (ML) is crucial to identify melanoma patterns. Therefore we need to organize the data so that we can apply ML on it. This paper presents a detailed characterization based on the most relevant knowledge in melanomas problem and how to relate them to apply Data Mining techniques to aid medical diagnosis in melanoma and improve the research in this field.

Keywords: Health Information Systems, Knowledge Management and Decision Support Systems, Melanoma domain, Computer Aided Systems.

1 Introduction

Melanoma is now one of the most social interesting cancers because it is more frequent in our society and affects people of any age. According to the *American Cancer Society* although is not the most common skin cancer, it is which causes most deaths. This increase, caused by solar habits, makes crucial the early diagnosis, even more if we analyze that this cancer is mortal in approximately twenty percent of cases [1,2] and prompt diagnosis permits practically a secure regain.

Nowadays domain experts works with some plain data bases with disjoined information that does not permit roomy experiments to identify melanoma patterns. In fact, these days diagnosis is not aided for computer systems and requires checking several reports, from different specialists, to give a unified prognostic. Dermatology experts from *Hospital Clinic i Provincial de Barcelona* (HCPB) want to do statistical analysis of its data, and characterize new melanoma patterns.

2

One way to improve the early diagnosis is to use Knowledge Management and Decision Support Systems based on the statistical results of the cancer information. This imply a Data Mining (DM) problem were we have to analyze high dimensional data, with uncertainty and missing values, to extract the patterns for aid decision making [3]. Some of these techniques have been proved in other kinds of cancer, like the breast one [4], with results that permits promising investigation. But this technology needs a good data characterization and organization to work and this is non-existent in its domain. To create it we have analyzed data from more than three thousand melanoma patients with information from reports of dermatologists, oncologists, surgeons, pathologists, and other specialists that work at HCPB. Some of these data are used in international studies that analyze specific aspects of the domain. But medical researchers want to study it on the whole to assess if clear patterns are reachable. This aim makes these data really unique and valuable for its variety and completeness.

Then, we have to propose a data organization in melanoma domain to find out which kind of information is relevant and how it is related. This proposal will be the base line for building a DM tool to help experts in melanoma research. Now we want to apply clustering techniques to divide the domain, as is intended previously with less data [5], and use Case-Based Reasoning to support the diagnosis.

The paper is organized as follows. Section 2 analyzes the framework in the field we would like to contribute. Section 3 describes the melanoma domain characterization. Finally, in sections 4 and 5 we discuss about the problem and mark the future work.

2 Related Work

Current works in the melanoma study treat the issue as a sectioned one. We could found works which use only some clinical and pathological data, but with a large number of cases [6], or that studies particular types of information over specific individuals [7,8]. This kind of reference is based on partial databases which are centred on concrete information that does not use the entire data of the domain. A support system that permits us to put it at stake could open new targets in the future using DM. Since early diagnosis of melanoma is the most important factor in the progress of the disease, the diagnostic accuracy is of major interest; is required to establish concise diagnostic criteria based on the clinical information. If we want to apply computer aided systems for getting these objectives, we need a complete and consistent relational model.

3 Melanoma Framework

To apply DM techniques in order to extract patterns from the complete data of the melanoma domain we have to unify the data from different plain databases, from

diverse medical specialities, medical information from non electronic support and new data from specific diagnostic machines. This section describes the complexity and constrains of melanoma domain. But, a basic conceptual explanation of the domain is not enough to represent all the relations, specific restrictions, and rules to follow in data insertion, maintenance, and search. Then, with all the collected information we have planed a relational model shown in Fig.1. This model permits all the storage capacities we need, respect the relations and prevent the insertion errors to avoid redundancies. Experts consider that the next aspects explain the domain: 1) person and family, 2) generic medical information, 3) tumors, 4) metastasis, and 5) controls and studies.

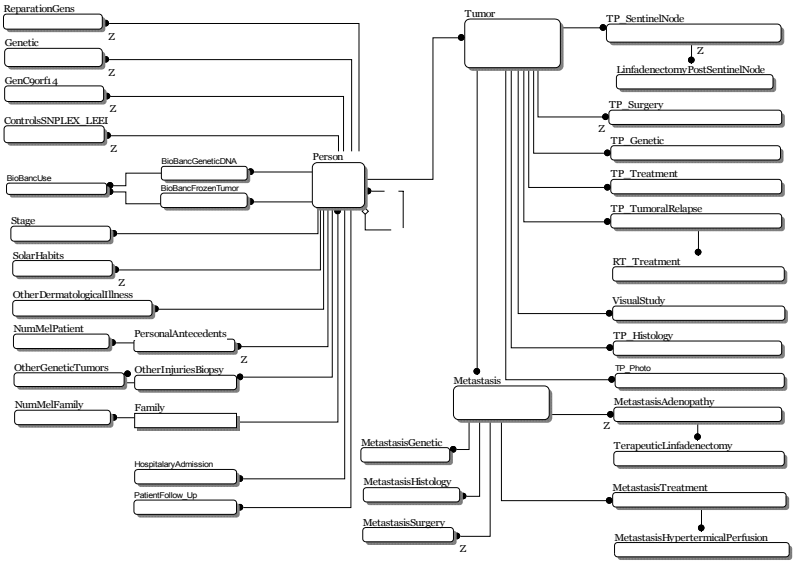

Fig. 1. Relational Model of Melanomas Domain.

Person and Family. This aspect summarizes the personal data and antecedents of a person. We note that there are different kinds of person: melanoma patients, familiars of these patients and control cases (healthy people). Patients are also related with their DNA samples and the hospitalary admissions and follow-up that permits to check the illness evolution. Emphasized information about the patient is to know the stage of the illness in different moments (indexed per date), the solar exposure, other dermatological illness, and its biopsies. Another important data to take into account is the familiar relations because it is possible a familiar development of the melanoma.

Tumors. This entity is related with a unique patient, but one patient could have more than one tumor. In addition, this entity has relation with its histologies, treatments applied, genetical information, surgery data, sentinel node, and metastasis. We also have the relation between the tumor and its different images.

4

Metastasis. This is a concept that depends on a tumor. Each tumor could have different metastasis. The general information is at the same time in relation with histologies, treatments, surgery, genetics, and lymph node metastasis that is included with the addition of linfadenectomy as a particular idea of a metastasis.

Controls and Studies. These aspects refer to the research studies developed by experts in the identification of patterns. A study is the analysis of a certain aspect of a patient. This concept is strongly linked with the person relation because they are who take part in the studies.

4 Discussion

The aim of our current research is to obtain clear melanoma patterns based on all the data existent in this domain. We would like to clusterize the complete information in order to create patterns of diverse characteristics that permit an easy prognostic from different kinds of melanoma even for non-expert doctors. We have also the target of obtaining a good ontological description that allows a better use of CBR techniques to aid medicians in diagnostic accuracy. But all this aims need a previous step that is the characterization of the domain that permits the desired investigation.

The characterisation exposed in section 3 structures melanoma domain and permits not only the target of allowing subsequent work of aid melanomas early diagnosis and accuracy, but it has a suitable insertion of information. This is important because we have found during the process that the previous data have inconsistencies or not related information that come from a bad pick up of it. The incorporation of triggers permits to control this situation and, in addition, filter the existent one in the migration process. Then, at the end of the process, we have obtained a new database completely consistent and with well related information.

We would like to note that the elaboration of the model adds new data obtained from paper reports and diverse databases not unified since now. This situation permits to prepare complete sets of raw data that allows the obtaining of melanoma patterns and statistics for medical research.

5 Conclusions and Further Work

Melanoma is a dermatological cancer with an important impact in our society. Medical researchers in this field want to use techniques to aid its diagnosis. The first step to permit the use of these techniques is the characterization of the domain in order to use the complete data in its studies. The definition proposed permits the creation of an application that allows: 1) An easy data introduction from the specialists, even during attendance work. 2) Knowing the trustworthiness grade of data in order to consider or not low reliable data in certain studies. 3) Rapid creation of new experiments and statistical results in medical research. 4) Study rela-

tions between different melanomas information in the bosom of a family. 5) Apply Data Mining Techniques in order to aid specialist to obtain optimal descriptions of patterns to improve the early diagnosis.

Nowadays, our work is focus in the final implementation of the application and the migration of the previous data. After this migration, we could apply Data Mining techniques to create a complete Knowledge Management and Decision Support System *ad hoc* to melanoma's domain that permits to help the medical researchers in the prognostic of this illness. This work are planed in two phases, the first one the extraction of patterns and explanations from this complete data, enlarging previous works that use parts of it [5]; and the second that is the implementation of a hierarchic reasoner that permits to aid the diagnosis by analysing the different kinds of information in separated stages to conclude an unified result.

Acknowledgements. We thank the Spanish Government for the support in MID-CBR project under grant TIN2006-15140-C03 and the Generalitat de Catalunya for the support under grants 2005SGR-302 and 2008FI_B 00499. We thank Enginyeria i Arquitectura La Salle of Ramon Llull University for the support to our research group. The work performed by S. Puig and J. Malvehy is partially supported by: Fondo de Investigaciones Sanitarias (FIS), grant 0019/03 and 06/0265; Network of Excellence, 018702 GenoMel from the CE.

References

1. Gutiérrez, R.M., Cortés, N.: Confronting melanoma in the 21st century. Med Cutan Iber Lat Am;35(1):3-13. (2007)
2. Tucker, M.A., Goldstein, A.M.: Melanoma etiology: where are we? Oncogene 22, 3042–3052. (2003)
3. Han, J., Kamber, M.: Data Mining: Concepts and Techniques. Elsevier. (2006)
4. Fornells, A., Golobardes, E., Bernadó E. & Martí, J.: Decision Support System for Breast Cancer Diagnosis by a Meta-Learning Approach based on Grammar Evolution. 9th ICEIS, pp. 222-227, INSTICC Press. (2006)
5. Fornells, A., Armengol, E., Golobardes, E., Puig, S. & Malvehy, J.:Experiences Using Clustering and Generalizations for Knowledge Discovery in Melanomas Domain. In 7th Industrial Conference on Data Mining, 2008, LNCS. 5077:57-71 Springer-Verlag. (2008)
6. Balch, C. M. et al.: Prognostic Factors Analysis of 17,600 Melanoma Patients: Validation of the American Joint Committee on Cancer Melanoma Staging System. J. Clin. Oncol., Vol 19: 3622-3634 – Am. Soc. Clin. Oncol. (2001)
7. Puig, S.: Role of the CDKN2A locus in patients with multiple primary melanomas. J. Clin. Oncol. 1;23(13):3043-51. (2005)
8. Malvehy, J. et al.: On behalf of the International Dermoscopy Society Board members. Dermoscopy report: Proposal for standardization Results of a consensus meeting of the International Dermoscopy Society. J. Am. Acad. Dermatol. 57(1):84-95. (2007)