User-adapted clinical decision support systems

Andrea Sboner *
Bioinformatics, ITC-irst
DIT, University of Trento
Via Sommarive, 18
38050 - Povo (Trento) I
sboner@itc.it

Riccardo Bellazzi
Dept. Inf. and System Sciences
University of Pavia
Via Ferrata, 1
27100 - Pavia, I
riccardo.bellazzi@unipv.it

Paolo Carli
Dept. of Dermatology
University of Florence
Via degli Alfani, 37
50121 - Florence, I
carli@unifi.it

Mario Cristofolini
Lega Italiana
per la lotta contro i tumori
Corso 3 Novembre, 134
38100 - Trento, I

Abstract

Clinical Decision Support Systems (CDSSs) have been one of the challenging real-world applications of artificial intelligence techniques for decades. CDSSs built so far mainly deal with explicit medical knowledge, typically encoded into a knowledge base. The rationale is that medical personal knowledge may be updated or time constraints limit physicians' ability to properly manage that knowledge. However, inter-user variability due to subjectivity in feature evaluation has been typically ignored when building clinical decision support systems. In some cases implicit medical knowledge can be important to share, especially for novices' education. In this work we present an approach for building CDSSs facing this problem through different artificial intelligence techniques: a machine learning approach for a critiquing module and a case based reasoning approach for a consulting module. The purpose is to build a system accounting for the specific skill an expertise of the user, providing new or enough information to take proper decisions.

Introduction

Clinical decision support systems (CDSS) have been one of the challenging real-world applications of artificial intelligence techniques for decades, since the pioneer studies in the sixties and early seventies (de Dombal *et al.* 1972; Shortliffe 1976; Miller, Jr, & Myers 1982).

CDSSs can be defined as: "active knowledge systems which use two or more items of patient data to generate case-specific advice" (Wyatt & Spiegelhalter 1991). Given this definition, CDSSs include, and exploit, three elements: a knowledge base, patient data, and an inference engine to provide case-specific advices. Within this framework, the user supplies input data (for instance patient data) to the system and the system provides its output by using the inference engine. The inference engine combines the data provided by the user and the data included into its knowledge base to produce its output. The knowledge base encodes the medical knowledge which is relevant for the system to solve the tasks it was built for.

The rationale for this framework is that physicians' knowledge can be incomplete in various ways: personal knowledge can be outdated or imprecise, medical knowledge contains uncertainties (due to the lack of definite biological understanding of diseases) that physicians are not able to properly manage, the development of biomedical equipment and instrumentation expanded investigative and therapeutic possibility of physicians, but overwhelmed them with new information that a busy physician is not able to handle, etc. Therefore, CDSSs can help them in managing this quantity of information in a better and more proper way to improve quality of their decisions.

Among the problems arising in the development of CDSS, the management of uncertainties in the medical knowledge has been deeply investigated, giving rise to various tools and methodological solutions. Those uncertainties have been typically included into the knowledge base through different approaches: probabilities in Bayesian systems, certainty factors in rule based system, to name a few.

However, most of these approaches deal with uncertainty in the explicit medical knowledge, getting rid of the possible difference in the subjective evaluation of patients' condition, i.e in the input data. Inter-user variability is a well known problem in medicine, caused by the lack of absolute reference values and by subjectivity. These are common factors in medicine, whenever no clear and definite knowledge of a disease is available (van Bemmel, Musen, & Miller 2000). Rather surprisingly, no solution is available in the literature to soundly cope with this problem (van der Lei 2000).

Subjectivity is a quite common characteristic in many decision processes. Large part of knowledge is not explicit, but tacit. Tacit knowledge is personal, context specific and thus hard to formalize and communicate. On the other hand, explicit knowledge can be easily communicated through a formal language for representation. Medical textbooks are examples of explicit knowledge; instead medical apprenticeship deals with tacit knowledge. As a matter of fact, medical training all over the world includes some form of apprenticeship that helps to teach the implicit medical knowledge to novices.

*Research topic is related to a PhD thesis in Information and Communication Technology at University of Trento, Italy. We would like to thank the dermatologists who, in different phases, participated in this study: A. Bergamo, P. Bauer (APSS, Trento,I), A. Chiarugi, P. Nardini, C.Salvini, M. Stante (Dept. of Dermatology, University of Florence, I).

In this work we describe a method for building clinical decision support systems accounting for the inter-user variability problem. In particular, we describe the creation of CDSSs able to adapt to the specific knowledge and skills of the user, by means of different artificial intelligence techniques. The medical domain of this work is the early diagnosis of melanoma, the most dangerous form of skin cancer.

The paper is organized as follows: next section describes the problem we faced. Following sections illustrate the solution we proposed and the experimental results we obtained. Finally a discussion on the results and on future developments is reported.

The problem, clinical background

The assessment of inter-user variability in medicine has been investigated a lot, mostly in the field of radiology and human pathology. The need for reducing as far as possible this problem especially in those disciplines which, in many cases, have a potential effects on diagnosis and treatment (histo-pathology for instance, provides a gold standard reference for the definition of many kinds of diseases) has led to protocols for quality control and to continual training sessions to ensure comparable levels of performances among physicians (e.g. in cytology, radiology, etc.).

Inter-user variability may deeply affect a CDSS relying on knowledge-base models. In fact, let us suppose we have a perfect knowledge of the patho-physiologic mechanism of a disease, i.e. we perfectly know how to diagnose a certain situation given a set of clinical findings. In this scenario a CDSS would apply its inference engine (activate a rule, compute a function, calculates probabilities, etc.) and provide the correct output given specific clinical findings. However, if those clinical findings are subjective, their assessment depends on the specific skill of the physician. If the degree of inter-user variability of those parameters is very high, the CDSS may fail in providing a correct assistance, even if "perfect" medical knowledge is available.

Early diagnosis of melanoma

Among skin tumors, malignant melanoma is the most dangerous as it is responsible of more than 75% of skin cancer deaths. Malignant melanoma develops from melanocytes, skin cells that produce the protective pigment melanin. The incidence of malignant melanoma has been constantly increasing worldwide during last century, especially for fair-skinned populations, at a rate of about 3-7% yearly. Fortunately, the mortality for melanoma increased at a lower rate than incidence. The main reason for this improvement in survival can be attributed to the earlier detection of melanoma, because there have been no major changes in melanoma treatment during the years.

Unfortunately, diagnosis of this kind of cancer is difficult because early lesions often have a benign appearance. The usual clinical practice of melanoma diagnosis is a visual inspection of the skin. In vivo epiluminescence microscopy (ELM, also called dermoscopy, dermatoscopy, or surface microscopy) has become an increasingly popular method of inspecting lesions (Pehamberger *et al.* 1993; Argenyi 1997). It is a non-invasive method that allows the investigators to look through the skin surface, making accessible structures that are beyond it. This feature greatly increases the morphological details that are visualized, providing additional diagnostic criteria to the dermatologist.

ELM inspection requires the assessments of specific features of skin lesions. Pattern analysis was one of the first methods developed for the diagnosis of pigmented skin lesions (PSLs), describing the entire set of lesion characteristics that has to be assessed (Pehamberger, Steiner, & Wolff 1987) and it is the most reliable method for teaching dermoscopy (Carli *et al.* 2003). Unfortunately, pattern analysis relies on features that can be difficult for non-expert to properly recognize. There is a lack of proper definition of these parameters, introducing subjectivity. Some studies highlighted that only well-trained dermatologists can benefit from dermoscopy, while non-expert physicians can even getting worse (Kittler *et al.* 2002; Binder *et al.* 1995). Moreover, a consensus among expert dermatologists on the dermoscopic features can help to improve the diagnosis suggesting that inter-rater variability is the main factor for the different performances of dermoscopy (Argenziano *et al.* 2003).

This situation is therefore particularly suited to investigate our approach dealing with inter-user variability.

User-adapted CDSS

The general architecture of the system is illustrated in figure 1.

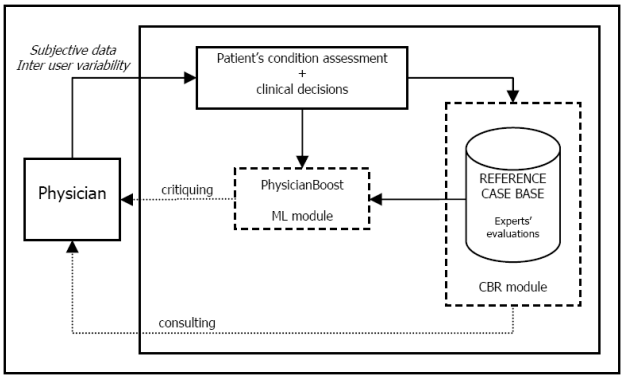

Figure 1: General architecture of the user-adapted CDSS

The CDSS is based on two modules: a critiquing module and a consulting module, both of them relying on a *reference dataset*. Such dataset is composed by a number of known cases. Objective information (such as patient's age, lesion location, etc.) and the "gold standard" data (e.g. the histological diagnosis) are present in the dataset. Expert physicians provide their subjective evaluations as well as their clinical decisions (diagnosis, treatment, etc.) independently. The reference dataset with the experts' evaluations and decisions is the *reference case base*.

Each new user of the CDSS is first required to evaluate the reference dataset by providing his/her clinical subjective evaluations and decisions. Such evaluations allow to:

- Build a critiquing module by applying a novel machine learning-based approach, able to use the evaluations of the single physician and to exploit his/her diagnostic capability. The critiquing system applies a classification model, learned by focusing on the most "difficult" cases of the reference dataset, i.e. the cases in which user's decision was wrong. The model adapts to the specific knowledge of the novice. This method is called "PhysicianBoost".

- Building a consulting module by comparing the user subjective evaluations with the ones contained in the reference case based. This allows: i) assessing similarities between the user and the experts in evaluating the lesions; ii) selecting the expert "more close" to the user in terms of the subjective evaluations; iii) assessing the diagnostic and treatment capabilities of the user with respect to the experts ones; iv) helping the critical revisions of user's subjective evaluations and clinical decision on previous cases. The consulting module aims to share the implicit knowledge of experts by means of a case based reasoning approach.

When a new case is provided by the user to the CDSS, the critiquing module would suggest the user to reevaluate his/her decisions if deemed necessary, i.e., when the clinical decisions and model outputs are different, while the consulting system shows experts' feature evaluations and decisions on similar cases in the case base upon user request: the retrieval takes into account the inter-user similarity estimated on the reference case base.

Reference Dataset

The reference dataset is composed by 177 pigmented skin lesions: 76 are malignant melanomas and 101 are benign lesions according to the histological diagnosis which provides the gold standard reference. Objective information, such as patient's age, lesion's location, etc., and digital images of clinical and dermoscopic analysis were collected during face-to-face visits at the Department of Dermatology, S.Chiara Hospital in Trento, Italy. All lesions were surgically excised and went through the histological examination to obtain the actual diagnosis.

Three expert dermatologists, with several years of experience in dermoscopy, evaluated each case in the reference dataset independently through a web based system. Objective information and a digital image set of the lesion were available to physicians so as to provide them with as much information as possible to perform subjective feature evaluations. In our experimental protocol we used the same dermoscopy parameters that are currently collected in the clinical practice in Trento (see table 1). In addition to subjective feature evaluation, experts also provided their clinical decisions in terms of clinical diagnosis (benign or malignant lesion) as well as the choice for excisional biopsy (yes or no). Three less-experienced dermatologists evaluated the reference dataset in the same way as experts. This procedure is the starting point for the development of the two modules of the CDSS.

The critiquing module

The critiquing module employs a machine learning algorithm to provide alerts to the user of the CDSS. A critiquing decision support system is supposed to provide unsolicited advices to the user. As an example, a critiquing system warns of possible patient's drug allergy when a physician prescribes that drug. In our case, system's alerts are tailored on the specific skill of that user.

The proposed method for the critiquing module, called PhysicianBoost, employs a machine learning algorithm, namely a Naive Bayes (NB). Typically, using machine learning algorithms to build CDSS requires the creation of a model from a set of known data, for instance patients' data whose final diagnosis is known. Then, the model is able to predict on unknown patient's data and it is normally supposed to be used by novices. Generally, the known data are provided by an expert or a group of experts. However, if input data are affected by inter-user variability, this model may fail when the system is used by novices.

Differently, in our approach, a model is created for each physician (a local model), starting from his/her feature evaluations. It then combines its output with clinical decision through a suitable combination rule, since the final aim is improving overall performances, i.e. the performances of physicians with the aid of the support system. The combination rule is domain dependent. The module accounts for both physician's assessment of feature values and his/her capability of correctly classify cases, which may turn out to be correlated. Figure 2 illustrates this approach from an abstract perspective.

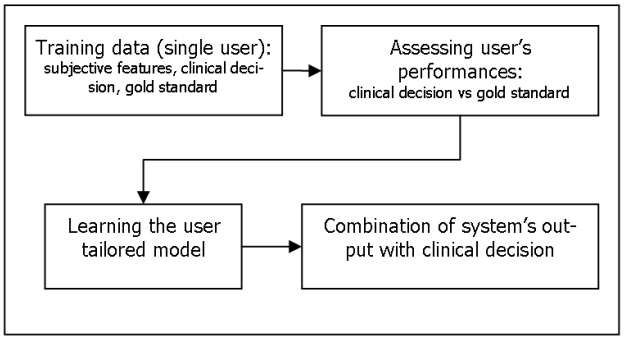

Figure 2: PhysicianBoost. First we assess physician's performances, then we build a personalized model through a machine learning algorithm. Finally we combine classifier output with clinical decisions

In more details, for each physician, a NB model was created, after placing more weight on physician's misclassified instances. In this way we focussed the NB model towards the most difficult cases for the physician. Afterward, clinical diagnosis and model output was properly combined trying to enhance sensitivity, the most critical parameter in this kind of medical application (Sboner *et al.* 2004). The combination rule yields "benign" only if both physician and model output are "benign"; otherwise is malignant. This combination rule is the basis for the creation of a critiquing interac-

tion modality with the system.

This choice resembles the real-world application of this critiquing system, in particular alerting the physician of a possible malignant lesion clinically diagnosed as benign.

The consulting module

The purpose of this module is to share experts' implicit knowledge contained into the subjective evaluations in the reference case base. The consulting module adopts case based reasoning paradigm to provide the user with expert knowledge. Case based reasoning is particularly suited for medical decision making when subjective knowledge is involved (Schmidt *et al.* 2001). Moreover, reasoning with cases corresponds with the typical decision making process of physicians; incorporating new cases in the reference case base updates the implicit knowledge. The consulting module is supposed to be activated by a specific request of the current user of the system. The scenario is therefore the following:

- the physician visits a new patient;
- he/she collects the objective data, if available and requested by the system;
- he/she provides the system with the subjective information as well as clinical decisions: diagnosis, treatment, etc.
- the physician then asks the system to seek for previously solved cases similar to that at hand;
- the system computes a similarity measure with the cases into the reference case base, ranking the cases accordingly;
- the physician can browse the ordered cases.

The ordered cases entail both the subjective feature evaluations as well as the clinical decisions of the experts. By allowing the comparison of those cases in terms of subjective feature evaluations and clinical decisions, this module supports the dissemination of expert tacit knowledge. The consulting module benefits from the multiple independent evaluations of the same cases given by the experts.

The most important point here is the retrieval of the similar cases. Again the inter-user variability prevents the naive application of search queries into the reference case-base. Therefore, two approaches were investigated to retrieve the similar cases:

1. a similarity measure is computed between the new case and those ones the physician previously evaluated which are included into his/her own reference dataset. The user can look at his/her "closer" examples, as well as directly compare experts' evaluations on those same cases;
2. the expert most similar to the single physician is assessed by means of Cohen's κ (see section Evaluation) computed on the subjective evaluations of the common reference case base. Once the "most similar" expert is identified, the system searches for the most similar cases in that expert's reference case base.

Despite which reference case base we search for the retrieval of cases, both approaches rely on a notion of similarity among cases. This similarity has to deal with numeric and nominal features as well as missing values. To account for all these aspects we defined a similarity among two cases u and v as:

$$sim(u,v) = 1 - \frac{1}{m}\sum_{i=1}^{m} d(u_i, v_i) \quad (1)$$

where $d(u_i, v_i)$ is a distance measure between the values u_i and v_i for the features i, and m is the number of the features. We employed a Manhattan distance for nominal features, whereas for ordinal and numeric features we used:

$$d(u_i, v_i) = \frac{|u_i - v_i|}{max(i) - min(i)} \quad (2)$$

where $max(i)$ and $min(i)$ are the maximum and minimum of attribute i, respectively. The normalization factor for numeric features ensures that $d(u_i, v_i) \in [0, 1]$.

To deal with missing values, we chose to assign $d(u_i, v_i) = 0.5$, if one of the parameter is missing. The rationale of this choice is to consider a missing data somehow between the complete agreement ($d(u_i, v_i) = 0$) and the complete disagreement ($d(u_i, v_i) = 1$). The distance measure is divided by the number of the features such that $sim(u, v) \in [0, 1]$.

Evaluation

According to the literature, we measured a great level of inter-user variability in the assessment of dermoscopic features. We measured the agreement by means of Cohen's κ defined as:

$$\kappa = \frac{p_{obs} - p_{exp}}{1 - p_{exp}} \quad (3)$$

where p_{obs} is the proportion of observed agreements and p_{exp} is the proportions of expected agreements. The possible values of κ range from -1 to 1. A κ equal to zero means that agreement is entirely attributable to chance; κ greater (less) than zero means that agreement is greater (less) than that expected only by chance. Generally, values of $\kappa < 0.40$ are considered poor agreement, $\kappa \in [0.40, 0.75]$ fair to good agreement, and $\kappa > 0.75$ excellent agreement.

Table 1 reports the average values of Cohen's κ for all the dermoscopic features and for the clinical decisions. Table 1 shows that there is high degree of inter-user variability, even if almost every feature is binary. However, despite the low level of agreement there is a substantial concordance in the clinical diagnosis suggesting that tacit, implicit knowledge is an important aspect in the early melanoma diagnosis.
For sake of simplicity, hereinafter we call physicians the less-experienced dermatologists in dermoscopy and experts the high-skilled dermatologists in dermoscopy.

Regarding the critiquing module, we compared the performances of our personalized method (PhysicianBoost) with those of physicians without aid and with those of a standard machine learning approach. For the latter case, as illustrated in figure 2, we first developed a "clean" dataset reducing inter-user variability. To achieve this, we combined

Table 1: Dermoscopy features and clinical decisions. Average agreement for all features among all six physicians.

Feature	**type**	**Cohen's κ**
Asymmetry	binary	0.34 ± 0.11
Border	binary	0.29 ± 0.12
Number of colors	ordinal [1,6]	0.38 ± 0.11
Atypical pigm. network	binary	0.36 ± 0.11
Abrupt network cut-off	binary	0.27 ± 0.13
Regression-Erythema	binary	0.47 ± 0.11
Hypo-pigmentation	binary	0.16 ± 0.12
Streaks	binary	0.48 ± 0.10
Slate-blue veil	binary	0.48 ± 0.10
Whitish veil	binary	0.17 ± 0.16
Globular elements	binary	0.35 ± 0.14
Comedo-like openings	binary	0.10 ± 0.19
Telangiectasia	binary	0.23 ± 0.17
Excisional biopsy	binary	0.38 ± 0.14
Clinical diagnosis	binary	0.63 ± 0.08

experts' evaluation on the reference case base with a majority vote simulating a "consensus" among experts. That "clean" dataset was used to create the NB model then used by the physicians.

Table 2 shows sensitivity and specificity of: physicians without aid, PhysicianBoost personalized system, and the standard approach, respectively. Those measure were estimated through a 10-fold cross validation. The algorithm was implemented in Java by means of the Weka library package (Witten & Frank 1999).

Table 2: Physicians' and PhysicianBoost performances in terms of sensitivity and specificity compared with the standard approach.

Physician	without aid		PhysicianBoost		standard	
	sens.	spec.	sens.	spec.	sens.	spec.
$Physician_1$	0.73	0.78	0.83	0.73	0.63	0.89
$Physician_2$	0.71	0.77	0.77	0.74	0.57	0.90
$Physician_3$	0.76	0.87	0.84	0.73	0.59	0.87

Our personalized critiquing system showed better performances in the diagnostic task with respect to the standard machine learning approach. The expert NB model, i.e. the model created on the clean dataset, has sensitivity and specificity of 0.76 and 0.84, respectively. The average sensitivity and specificity among experts are 0.88 and 0.75, respectively.

Table 2 shows that our approach, PhysicianBoost, is able to improve the sensitivity of the physicians. PhysicianBoost's specificity decreases, however it is still comparable to experts average specificity (0.75). As an example, for $Physician_1$, the sensitivity increases of 10%. This means that the critiquing module helps recognizing 8 malignant lesions previously misdiagnosed, while making only 5 false positives more. On the contrary the standard approach has very low sensitivity values. Those values are even lower than those of the physicians without aid.

Concerning the consulting system, we computed the similarity for each case in the physician's reference dataset following the two approaches. Each case in turn was considered as a new case and the similarity was computed on the remaining cases. Table 3 shows the average similarity we obtained for the top-ranked case.

Table 3: Overall similarity values for the two approaches regarding the top-ranked case. Mean values and ranges [min-max] are shown.

	1^{st} approach	2^{nd} approach
$Physician_1$	0.85 [0.74-0.96]	0.83 [0.73-0.94]
$Physician_2$	0.85 [0.73-0.98]	0.82 [0.71-0.95]
$Physician_3$	0.84 [0.72-0.97]	0.84 [0.68-0.97]

The similarity values for the two approaches are quite high, however it is worth noting that the actual relevance of the retrieved case should be assessed by the physician. Assessing if the retrieved cases are relevant to the problem at hand is demanded to the physician because the complete information is contained in the images, while we rely only on a coded description. This aspect is a common problem in CBR systems in medicine (Schmidt *et al.* 2001).

However, despite how relevant the top-ranked case is both approaches allows some consideration. Let us first assume that the top-ranked case is relevant, thus:

- 1^{st} approach. Physicians can compare their own subjective evaluations and clinical decisions on the similar cases in their own reference case base. In addition, they can assess the differences with experts by looking at the same cases into experts' reference case base. This is possible because the reference dataset is common to physicians and experts.
- 2^{nd} approach. The physician can compare directly expert's clinical decisions on cases whose feature evaluation is most similar. In this case, if the clinical decisions are the same of the physician, he/she can be reassured about the correct clinical reasoning. On the other hand, if the expert conclusions were different, physicians can figure out why there is such difference.

On the other hand, if the retrieved case is not relevant, the user can figure out the reason for that. In particular we have these two situations, depending on the approach:

- 1^{st} approach. This means that the physician provided similar evaluations to cases judged not relevant to the problem at hand. A possible result would be the re-vision of the evaluations, with the help of corresponding experts information;
- 2^{nd} approach. This means that the expert evaluated the case similarly to the physician. The physician can critically assess the proper evaluations accordingly to the low actual relevance with expert's case.

Discussion

CDSS built so far are dealing with explicit medical knowledge. For instance, the new clinical paradigm of Evi-

dence Based Medicine (EBM) aims at providing physicians with scientifically grounded "best practice" for care delivery. This paradigm shift in clinical care has led to the development of CDSSs based on computerized guidelines, benefitting from the great deal of work carried out in the field of knowledge representation in medical informatics.

However, in some cases inter-user variability may deeply affect the performances of a knowledge based CDSS. This problem, despite being well known in medicine, has been typically ignored when creating CDSSs, leading in some cases to inadequate CDSSs for novices, as noted by Cross *et al.*

In our approach, regarding the early diagnosis of melanoma, we proposed the development of two modules to tackle inter-user variability problem: a critiquing module based on a novel machine learning approach, which adapt to user's expertise; and a consulting module, which support the sharing of experts' implicit knowledge.

Regarding the critiquing system, our solution tackles both the inter-user variability problem as well as the provision of user-adapted advices. PhysicianBoost performances, compared both with unaided physicians' ones and with those of standard solutions, showed that inter-user variability is properly addressed, provided the limited data at our disposal. In fact, the results we obtained showed that our approach (PhysicianBoost) is able to improve the diagnostic performances of physicians in terms of sensitivity, i.e. the recognition of the malignant lesions. On the other hand, though decreasing, specificity is comparable to that of expert dermatologists. It is worth noting that while missing a melanoma is potentially life-threatening, misdiagnosing a benign lesion means an unnecessary surgery, generally carried out in an outpatient clinic.

The Naïve Bayes classifier can also partially account for the improvement of physician's expertise by incrementally updating the probabilities in each physician's model, as long as new cases are inserted into the system. Moreover, the consulting module of the CDSS may induce the user to re-evaluate some cases in the reference dataset, accordingly to the new acquired skills. Again, in this case, the Naïve Bayes model is easily updated.

In addition, the consulting module of the CDSS provides physicians with insights into personal, tacit knowledge of experts, as well as a method for revising their personal knowledge. Similarly to other authors (Bozec *et al.* 1998), we exploit a case retrieval based on a coded description of images. In our application, a content-based image retrieval would be an alternative solution. However, this approach is not feasible because images in the reference case base were acquired with different acquisition devices leading to very diverse images from a computerized image processing perspective. In addition, this choice also allows a physician without digital image acquisition devices to use this CDSS. These aspects would foster the actual implementation and use of the system in the clinical practice. Nevertheless, the ultimate assessment should evaluate how clinical decisions are affected by the consulting systems in a long-term training framework.

As future development, both the consulting module and the critiquing one should surely benefit from the increase in the number of cases into the case base. This can ensure a proper coverage of pigmented skin lesions domain, and possibly include other kinds of dermatologic lesions (such as seborrheic keratosis) as well as other relevant clinical information (such as the "ugly duckling" sign).

References

Argenyi, Z. 1997. Dermoscopy (epiluminescence microscopy) of pig-mented skin lesions. current status and evolving trends. *Dermatol Clin* 15(1):79–95.

Argenziano, G.; Soyer, H.; Chimenti, S.; Talamini, R.; Corona, R.; Sera, F.; Binder, M.; Cerroni, L.; and et al. 2003. Dermoscopy of pigmented skin lesions: results of a consensus meeting via the internet. *J Am Acad Dermatol* 48(5):679–693.

Binder, M.; Schwarz, M.; Winkler, A.; Steiner, A.; Kaider, A.; Wolff, K.; and Pehamberger, H. 1995. Epiluminescence microscopy. a use-ful tool for the diagnosis of pigmented skin lesions for formally trained dermatologists. *Arch Dermatol* 131(3):286–291.

Bozec, C. L.; Jaulent, M.; Zapletal, E.; Heudes, D.; and Degoulet, P. 1998. Idem: a web application of case-based reasoning in histopathology. *Comput Biol Med* 28(5)(5):473–487.

Carli, P.; Quercioli, E.; Sestini, S.; Stante, M.; Ricci, L.; Brunasso, G.; and De Giorgi, V. 2003. Pattern analysis, not simplified algorithms, is the most reliable method for teaching dermoscopy for melanoma diagnosis to residents in dermatology. *Br J Dermatol* 148(5):981–984.

Cross, S.; Stephenson, T.; Mohammed, T.; and Harrisont, R. 2000. Validation of a decision support system for the cytodiagnosis of fine needle aspirates of the breast using a prospectively collected dataset from multiple observers in a working clinical environment. *Cytopathology 11(6):503-12, 2000* 11(6):503–512.

de Dombal, F.; Leaper, D.; Staniland, J.; McCann, A.; and Horrocks., J. 1972. Computer-aided diagnosis of acute abdominal pain. *Brit Med Jour* 1:376–380.

Kittler, H.; Pehamberger, H.; Wolff, K.; and Binder, M. 2002. Diagnostic accuracy of dermoscopy. *Lancet Oncol 3(3):159-65, 2002.* 3(3):159–165.

Miller, R.; Jr, H. P.; and Myers, J. 1982. Internist-1, an experimental computer-based diagnostic consultant for general internal medicine. *N Engl J Med* 307:468–476.

Pehamberger, H.; Binder, M.; Steiner, A.; and Wolff., K. 1993. In vivo epiluminescence microscopy: improvement of early diagnosis of melanoma. *J Invest Dermatol* 100:356S–362S.

Pehamberger, H.; Steiner, A.; and Wolff, K. 1987. In vivo epiluminescence microscopy of pigmented skin lesions. i. pattern analysis of pigmented skin lesions. *J Am Acad Dermatol* 17(4):571–583.

Sboner, A.; Azzini, I.; Demichels, F.; Bauer, P.; Bellazzi, R.; Carli, P.; and Cristofolini, M. 2004. A combined human-computer method for building effective decision

support system. In Zupan, B., and Holmes, J., eds., *Working notes of the 9th IDAMAP Workshop - Intelligent Data Analysis in Medicine and Pharmacology*.

Schmidt, R.; Montani, S.; Bellazzi, R.; Portinale, L.; and Gierl, L. 2001. Case-based reasoning for medical knowledge based systems. *Int J Med Inform* 64:355–367.

Shortliffe, E. 1976. *Computer-Based Medical Consultation: MYCIN*. New York: Elsevier/North Holland.

van Bemmel, J.; Musen, M.; and Miller, R. 2000. *Handbook of Medical Informatics*. Heidelberg: Springer-Verlag, 2nd edition. chapter Methods for decision support.

van der Lei, J. 2000. *Handbook of Medical Informatics*. Heidelberg: Springer-Verlag, 2nd edition. chapter Clinical decision-support system.

Witten, I. H., and Frank, E. 1999. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann.

Wyatt, J., and Spiegelhalter, D. 1991. Field trials of medical decision-aids: potential problems and solutions. In *Proc Annu Symp Comput Appl Med Care*, 3–7.