A Medical, Description Logic based, Ontology for Skin Lesion Images

Manolis Maragoudakis, Ilias Maglogiannis *Member IEEE*, and Dimitrios Lymberopoulos, *Member IEEE*

***Abstract*— Researchers have portrayed an increasing effort towards providing formal computational frameworks to consolidate the plethora of notions and relations used in the medical domain. Despite the fact that there are many reasons for this, the need for standardization of protocols and terminology is critical, not only for the provision of uniform levels of health care, but also to facilitate medical science research. In the domain of skin lesions, the variability of semantic features contained within images is major among the barriers to the medical understanding of the symptoms and development of early skin cancers. Such variability is acknowledged across specialist fields of medicine, motivating standardization of terminologies for reporting medical practice. The desideratum of making these standards machine-readable has led to their formalization in the form of ontologies. Ontologies are computational artifacts designed to provide semantic representations of a particular domain of interest. Such a representation can be encoded and reused, allowing navigation of the key concepts recorded and retrieval of information indexed against it. This fact bridges the required standardization gap by offering a set of labeling options to record observations and events encountered by medical experts. Given the twin goals of ontologies -- representation and standardization – the present paper deals with the creation of an ontology for skin lesion images, encoded using Description Logic in OWL format, which can be used for decision support systems that operate on a classification basis. The declarative framework within which ontologies are encoded can transcend any specific application context and help dermatologists perform semantic inference on skin lesion images. Furthermore, extensions are easier to be accomplished and new classes can be straightforwardly incorporated.**

I. Introduction

THERE are increasing efforts towards formal frameworks and protocols for notions of the medical domain. Simultaneously, a major challenge for the Web is to evolve towards its semantic evolution, named as «*Semantic Web*», in which information is possible to be tagged with semantics, enabling human and machines to make a better use of information, and achieve accurate inference on future, previously unseen data. The semantic markup of resources is a means to reach this goal. To standardize a semantic markup method, the Semantic Web proposes to rely on the one hand on a uniform formalism, e.g. XML or OWL, and on the other hand on an organization of knowledge into ontologies. Ontologies play a central role in the Semantic Web, since they define a precise and shared vocabulary for the semantic markup of the resources and their description by metadata. Ontologies are the key technology to explicitly describe the semantics of the information and enabling to exchange contents.

Given the fact that advances in medical technologies produce a vast amount of textual and non-textual data (e.g. images), annotation of especially the latter is very important for medical decision support systems since medical experts have difficulties in understanding the criteria of decision in many existing systems. A semantic classification and hierarchy of a medical application using ontologies would be beneficial when rather general or ambiguous queries are posed by users. The search and potential-solution spaces make it increasingly problematic for medical experts to locate useful information without semantic categorization.

In this paper, we examine the application and development of ontologies in real world domains. A bottom-up approach inevitably introduces a complication: any real-world application of ontologies will be relative to some task or group of tasks. However, as we shall show, the issues raised by task relativity are very strongly related to the issues raised by the question of usability in general, and these, we contend, cannot be ignored.

More specifically, we introduce our remarks of constructing a visual ontology for the domain of skin lesions, by accommodating image experimental instances and extracting features that define semantic concepts and inter-relations. As with other, similar approaches for different medical domains, our approach is based on the assumption that: given the skin lesion images, (i) expressing a the necessary feature space using domain knowledge is feasible; (ii) the marking up and the annotation of the regions of interest is practical; and (iii) representing and reasoning about the textual descriptions is able to be performed with a reasonable complexity in a given context of user queries.

We are motivated by the fact that diagnosis of skin lesions in our days is mostly based on the visual assessment of pathological skin and the evaluation of dermatological macroscopic features by the medical professionals. This issue shows the high dependency of the correct diagnosis with the observer's experience as well his visual acuity. However, the interest of the biomedical scientific

Manolis Maragoudakis is with the University of the Aegean, Department of Information and Communication Systems Engineering, Samos, Greece, phone: +30 22730 82000, fax: +30 22730 82001, e-mail: mmarag@aegean.gr.

Ilias Maglogiannis is with the with the University of the Aegean and the University of Central Greece, Lamia, Greece, (e-mail: imaglo@aegean.gr).

Dimitrios Lymperopoulos is with the Electrical Engineering Department, University of Patras, Patras, Greece, (e-mail: d.lympero@wcl.ee.upatras.gr).

community for computer supported skin lesion inspection and characterization has been increased during the last years [1]-[7]. The analysis of digital images containing pigmented skin lesions and the quantification of tissue lesion features has been proved to be of essential importance in clinical practice, because several tissue lesions can be identified based on measurable features extracted from an image. In addition the use of digital image features may help in an objective follow up study of skin lesion progression and test the efficacy of therapeutic procedures.

Skin cancer is among the most frequent types of cancer and one of the most malignant tumors. Its incidence has increased faster than that of almost all other cancers and the annual rates have increased on the order of 3–7% in fair-skinned populations in recent decades [8]. The cutaneous melanoma, which is the most common type of skin cancer, is still incurable. However when it is diagnosed at early stages it can be treated and cured without complications. The differentiation of early melanoma from other pigmented skin lesions (e.g., benign neoplasms that simulate melanoma) is not trivial even for experienced dermatologists; in several cases primary care physicians seems to underestimate melanoma in its early stage [9].

The creation of the ontology for skin lesions was made in OWL format, which benefits from the domain of Description Logics (DL), in order to establish a uniform syntax, free of semantic ambiguity and feasible to be accompanied by a strong inference engine. The advantages of a DL-based approach are especially evident when considering medical images whose interpretation involves substantial domain knowledge.

The paper is structured as follows: section 2 sketches the domain of skin lesion. Section 3 introduces the extracted image features that define classes and properties for the ontology preparation. Section 4 discusses the visual ontology for skin lesion images and inference potential, facilitated by OWL.

II. Skin Cancer Background Information

The skin consists of a number of layers with distinct function and distinct optical properties. White light shone onto the skin penetrates superficial skin layers and whilst some of it is absorbed, much is remitted back and can be registered by a digital camera. The stratum corneum is a protective layer consisting of keratin-impregnated cells and it varies considerably in thickness. Apart from scattering the light, it is optically neutral. The epidermis is largely composed of connective tissue. It also contains the melanin producing cells, the melanocytes, and their product, melanin. Melanin is a pigment which strongly absorbs light in the blue part of the visible and in the ultraviolet spectrum. In this way it acts as a filter which protects the deeper layers of the skin from harmful effects of UV radiation. Within the epidermal layer there is very little scattering, with the small amount that occurs being forward directed. The result is that all light not absorbed by melanin can be considered to pass into the dermis. The dermis is made of collagen fibers and, in contrast to the epidermis, it contains sensors, receptors, blood vessels and nerve ends (see Figure 1).

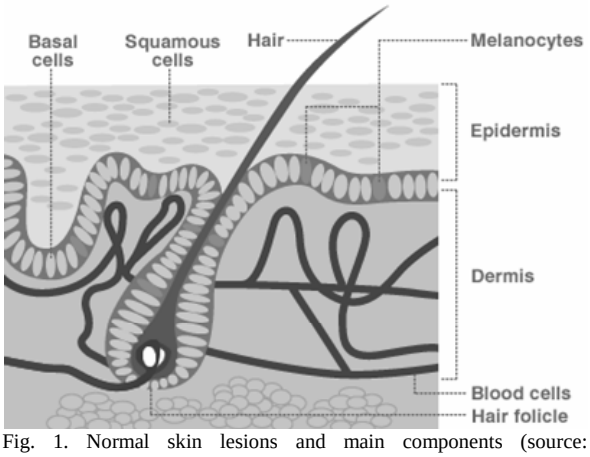


Fig. 1. Normal skin lesions and main components (source: MediceNet)

Pigmented skin lesions appear as patches of darker color on the skin. In most cases the cause is excessive melanin concentration in the skin. In benign lesions (e.g. common nevi) melanin deposits are normally found in the epidermis (see Figure 3(c)). In malignant lesions (i.e. melanoma), the melanocytes reproduce melanin at a high, abnormal rate (see Figure 2). Whilst they and their associated melanin remain in the epidermis, melanoma is termed ‘in situ’. At this stage it is not life-threatening and its optical properties make it conform to those of the normal, highly pigmented skin. When malignant melanocytes have penetrated into the dermis, they leave melanin deposits there, changing the nature of skin coloration. The presence of melanin in the dermis is the most significant sign of melanoma. However, it cannot be used as a sole diagnostic criterion because in situ

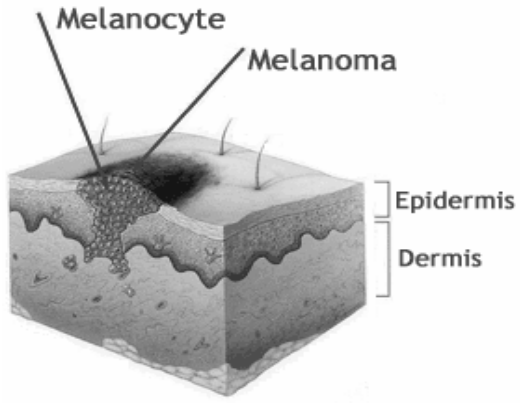
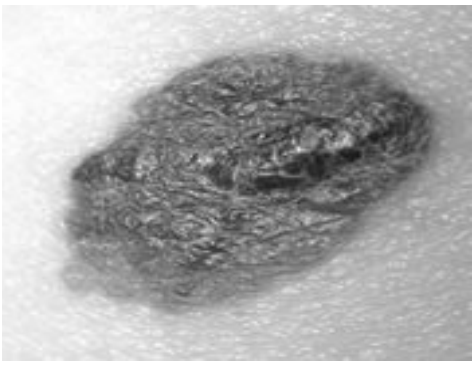


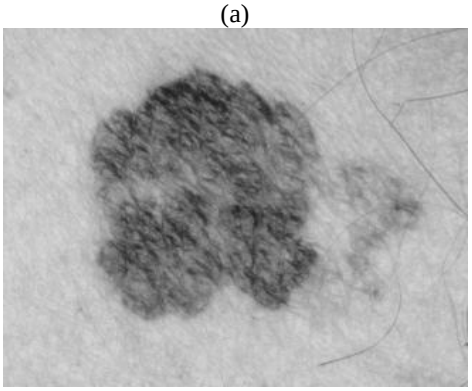
Fig. 2. Illustration of Melanocytes and Melanoma on skin (source: MediceNet)

melanomas do not have dermal melanin. Moreover, some benign nevi have dermal deposits, although their spatial patterns tend to be more regular than in melanoma. Other signs, some of which can be indicative of melanoma in situ,

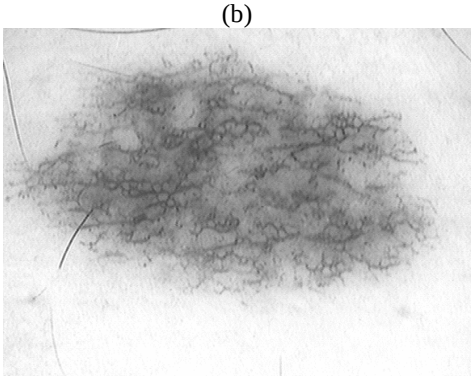
are thickening of the collagen fibers in the papillary dermis (fibrosis); increased blood supply at the lesion periphery (erythematic reaction); and lack of blood within the lesion, in the areas destroyed by cancer. The colors associated with skin which has melanin deposits in the dermis normally show characteristic hues not found in any other skin conditions. This provides an important diagnostic cue for a clinician. If the visual approach corroborates a suspicion of skin cancer, histology is needed to make explicit diagnosis. Figure 3 illustrates typical example skin lesions of melanoma, dysplastic (benign) nevus and non-dysplastic (common) nevus.



(a)



(b)



(c)

Fig. 3. Images of (a) typical melanoma, (b) dysplastic nevus and (c) non-dysplastic (common) nevus

III. Image Analysis and Ontology Pre-Processing

In automated diagnosis of skin lesions, feature design is based on the so-called ABCD-rule of dermatology. ABCD rule, which constitutes the basis for a diagnosis by a dermatologist [6] represents the *Asymmetry*, *Border* structure, variegated *Color*, and the *Differential Structures* characteristics of the skin lesion. The feature extraction is performed by measurements on the pixels that represent a segmented object allowing non-visible features to be computed. Several studies have also proven the efficiency of border shape descriptors for the detection of malignant melanoma on both clinical and computer based evaluation methods [7], [8]. These features may be used for the semantic annotation of dermatological digital images allowing for a quantitative and qualitative description of the image content, useful for diagnostic or training purposes. Three types of features are analyzed in this study: Border Features which cover the A and B parts of the ABCD-rule of dermatology, Color Features which correspond to the C rules and Textural Features, which are based on D rules. The aforementioned grouping of features was used for building the corresponding ontological model.

More specifically the utilized in this study features are grouped as follows:

A. Border Features

1. Thinness Ratio measures the circularity of the skin lesion defined as

$$TR = 4\pi Area/(Perimeter)^2$$

2. Border Asymmetry is computed as the percent of non-overlapping area after a hypothetical folding of the border around the greatest diameter or the maximum symmetry diameters
3. The variance of the distance of the border lesion points from the centroid location.
4. Minimum, maximum, average and variance responses of the gradient operator, applied on the intensity image along the lesion border.

B. Color Features

1. Plain RGB color plane average and variance responses for pixels within the lesion
2. Intensity, Hue, Saturation Colour Space average and variance responses for pixels within the lesion

$$I = \frac{R+G+B}{3}$$

$$S = 1 - \frac{3}{R+G+B}[\min(R,G,B)]$$

$$W = \arccos\{\frac{R - \frac{1}{2}(G+B)}{[(R-G)^2 + (R-B)(G-B)]^{1/2}}\}$$

where:

$$H = W \qquad if\ G>B,$$

$$H = 2\pi - W \qquad if\ G<B$$

$$H = 0 \qquad if\ G=B$$

3. Spherical coordinates LAB average and variance responses for pixels within the lesion

$$L = \sqrt{R^2 + G^2 + B^2}$$

$$AngleA = \cos^{-1}[\frac{B}{L}]$$

$$AngleB = \cos^{-1}[\frac{R}{L\sin(AngleA)}]$$

C. Color Features

1. Dissimilarity, *d*, which is a measure related to contrast using linear increase of weights as one moves away from the GLCM diagonal.

$$d = \sum_{i,j=0}^{N-1} P_{i,j}|i-j|$$

where *i* is the row number, *j* is the column number, *N* is the total number of rows and columns of the GLCM matrix, and

$$P_{i,j} = \frac{V_{i,j}}{\sum_{i,j=0}^{N-1} V_{i,j}}$$

is the normalization equation in which $V_{i,j}$ is the DN value of the cell *i,j* in the image window.

2. Angular Second Moment, *ASM*, which is a measure related to orderliness, where $P_{i,j}$ is used as a weight to itself :

$$ASM = \sum_{i,j=0}^{N-1} P^2_{i,j}$$

3. GLCM Mean, μ_i , which differs from the familiar mean equation in the sense that it denotes the frequency of the occurrence of one pixel value in combination with a certain neighbor pixel value and is given by

$$\mu_i = \sum_{i,j=0}^{N-1} i(P_{i,j}).$$

For the symmetrical GLCM, $\mu_i = \mu_j$.

4. GLCM Standard Deviation, σ_i, which gives a measure of the dispersion of the values around the mean

$$\sigma_i = \sqrt{\sum_{i,j=0}^{N-1} P_{i,j}(i-\mu_i)^2}$$

IV. Skin Lesion Imaging Ontologies

Dermatology as a medical specialty is a very complex domain for modeling and representing intended meaning. Hence building domain ontology from scratch is a complicated task and it requires a general, foundational ontology to interoperate and support soft modularization. The foundational ontology implements the most appropriate set of principles and speed up the ontology building process as "reinventing the wheel" is avoided. Therefore, the proposed approach includes a visual ontology providing the concepts and properties for indexing the visual content of the images as well as an instance dataset of various skin lesions.

The visual and domain ontologies are being developed using the Web Ontology Language OWL, which according to [13] is the recommended standard for ontologies. Interoperability is our main motivation for using OWL. Furthermore, the expressiveness which is achieved with OWL and the precise formal semantics were also of significant interest for our approach. Using DL and based on subsumption calculus, OWL supports ontology automatic classification and consistency checking and identification of the class in instances. For our experiments, we utilized the Racer (www.racer-systems.com) inference engine.

In order for the ontology to be expandable and taking into consideration that for the time being, no standard for skin lesion images is proposed, our construction strategy follows a pyramid model, illustrated in Figure 4. Our model is a multi-level hierarchy in which a set of graphical descriptors contains a set of image subjects and subsequently, this contains a set of medical terminology. For example, as shown in Figure 5, the descriptive features of a surrounded region corresponding to the concept Region-Of-Interest (RoI) are represented using descriptors such as Border, Color and Texture. Subsequently, an instance of RoI can be used to construct other abstract objects, e.g. the instances of Medical-Image, by referring to the instances of RoI via the

property *contains*.

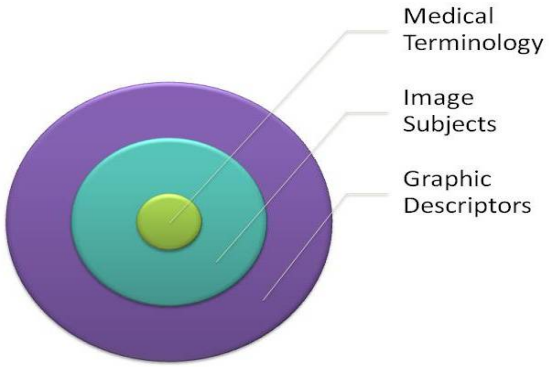


Fig. 4. Hierarchical Model of the Skin Lesion Visual Ontology

The knowledge taxonomy was originated by the image feature descriptors presented in Section 3. The class hierarchy as encoded in OWL format is depicted in Figure 6.

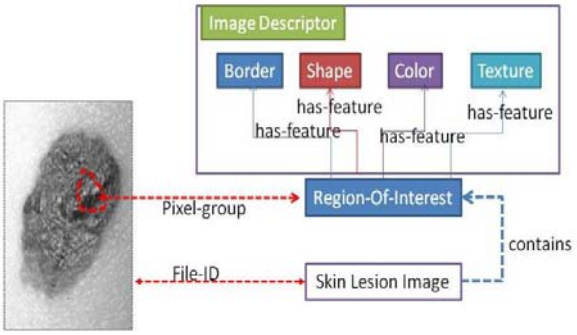


Fig. 5. Characterizing a RoI for a skin lesion image sample.

The visual ontology was enriched with attributes, domain and range constraints using DL, in order to allow for better annotation of skin lesion image instances. It is agreed that selecting a DL language with the right expressive power is crucial as a "trade-o " between the language expressiveness and the computational complexity is inevitable [14]. As regards to our ontology, the *SHIQ* DL [15] was considered as the most suitable modeling language under both theoretical and practical considerations. On one hand, the characteristics of the imaging ontology suggest that the selected modeling language should be equipped with the expressive and deductive power for qualified role value restrictions (both existential and universal), role hierarchies, inverse roles and qualified role number restrictions, features of which are fully covered by the *SHIQ* language. We utilized the RACER inference engine, so that users could perform reasoning for new instances in an acceptable computational cost. Furthermore, General Concept Inclusion (GCI) was ruled out—only simple concept names are allowed as the left hand operands in a concept introduction axiom—in an attempt to reduce the computational complexity it causes [16]. In our system, both image annotation and retrieval are supervised by the aforementioned DL-based inferential engine.

Medical experts are likely to consider constructing ontologies difficult, especially, the use of existential ( ) and

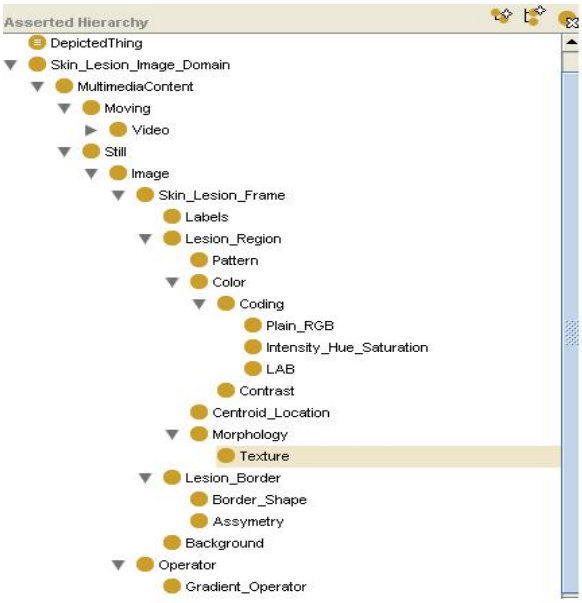
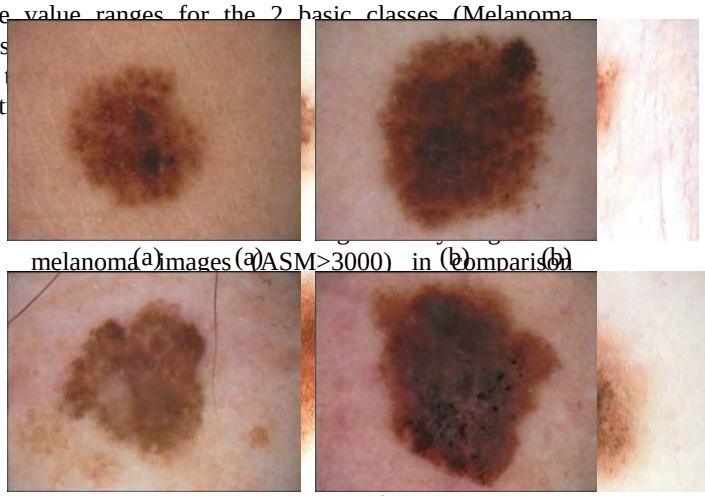


Fig. 6. Class taxonomy screenshot

universal ( ) quantifiers. As a result, the task of modifying the conceptual structure of ontologies is normally left to knowledge engineers. Furthermore, we do not expect the medical experts to be keen and active in extending the ontology axioms. However, with the aid of Data Mining techniques and by extending association rules to ontology properties, medical experts can benefit by facilitating DL-based classification and querying on database instances [11]. The following examples demonstrate the phase of translating raw data describing skin lesion images to basic DL descriptors. The examples contain images accompanied with features subset from Melanoma lesions (Figure 7) and Displastic Nevus lesions (Figure 8). Some remarks on the feature value ranges for the 2 basic classes (Melanoma lesions



apply t
except

- Melanoma lesions are also more concentrated around the center, since the variance of the distance of the
- melanoma images (ASM>3000) in comparison
- border lesion which is lower for non-malignant cases (H-std<8).

NAME	Mean R	NAME	ASM Mean R	H std	ASM std G	H std	Distance std G	Distance std	Assyme-try
a	122,74		4324,26	14,08	17,13		11,4	17,49	54,55
b	91,11	b	15253,3 4	159,81	2197,53	5,4	28,09	31,58	31
c	87,94	c	17524,5 3	37,94	82,17	11,23	9,72	14,91	10,91 53,78
d	135,14	d	3008,79	69,94	16,17		20,08	12,42	9,36 88,21

Fig. 8. Displastic Nevus lesions images and data

border lesion points from the centroid location is lower (distance-standard <20). The corresponding feature for Dysplastic Nevus Lesions is >25 in most of the cases.

The above remarks could be translated to the following DL statements:

- RoI_Dysplastic_Nevus=… hasmorphological_features.Color.Coding.RGB.hasErythema.hasMeanR≥*n*…∪…=… hasmorphological_features.Texture.hasASM<*m*… hasmorphological_features.Border.Shape.hasCentroid_Distance≥*k*.…=… hasmorphological_features.Color.Coding.Intensity_Hue_Saturation.hasHue≤*h*…
  - (n≈140, m≈3000, k≈25, h≈8)
- RoI_Melanoma=… hasmorphological_features.Color.Coding.RGB.hasErythema.hasMeanR≤*n'*…∪…=… hasmorphological_features.Texture.hasASM>*m'*… hasmorphological_features.Border.Shape.hasCentroid_Distance<*k'*.…=… hasmorphological_features.Color.Coding.Intensity_Hue_Saturation.hasHue≥*h'*…
  - (n'≈150, m'≈3000, k'≈20, h'≈14)

Researchers could use the above DL in order to perform a statistical-based evaluation on the numeric data and validate the classification ability of the ontology engineering structure. The DL ensures that semantic relationships are taken into account during comparison of classified examples using quantity (data) and quality (classes) methods.

REFERENCES

[1] S. Tomatis, C Bartol, G. Tragni, B. Farina, R. Marchesini : "Image analysis in the RGB and HS colour planes for a computer assisted diagnosis of cutaneous pigmented lesions" Tumori vol 84 pp 29-32 1998.

[2] J. Sanders B. Goldstein, D. Leotta K. Richards : "Image proccesing techniques for quantitative analysis of skin structures" Computer Methods and Programs in Biomedicine 59 pp 167-180 1999.

[3] G. Hansen, E. Sparrow, J. Kokate, K. Leland, P. Iaizzo : "Wound Status Evaluation Using Color Image Processing" IEEE Transactions on Medical Imaging, vol16, no1 pp 78-86 Feb 1997.

[4] M. Herbin, F. Bon, A. Venot, F. Jeanlouis, M. Dubertret, L. Dubertret, G. Strauch : "Assessment of Healing Kinetics Through True Color Image Processing" IEEE Transactions on Medical Imaging, vol12, no1 pp 39-43 Mar 1993.

[5] S. Chin : "The assessment of methods of measurements" Stat. Med. Vol.9 pp 351-362, 1990.

[6] W. Lohman, E. Paul : "In situ detection of melanomas by fluorescence measurements" Naturewissenschaften 1988, 75 201-202.

[7] Bono, S. Tomatis, C. Bartoli : "The invisible colors of melanoma. A telespectrophotometric diagnostic approach on pigmented skin lesions" European Journal of Cancer 1996 32A, 727-729.

[8] Marks R. Epidemiology of melanoma. Clin Exp Dermatol 2000, 25:459–63.

[9] Pariser R.J. and Pariser D.M., "Primary care physicians errors in handling cutaneous disorders", J Am Acad Dermatol, 17, pp. 239-245, 1987.

[10] R. Stevens, C. Wroe, S. Bechhofer, P. Lord, A. Rector, andC.Goble.Building ontologies inDAML+OIL. Comparative and Functional Genomics, 4(1), 2003.

[11] T. Bittner and S.Winter. On ontology in image analysis. Lecture Notes in Computer Science, 1737:168–191, 1999.

[12] T. Berners-Lee, J. Hendler, and O. Lassila. The SemanticWeb. Scientific American, pages 28–37,May 2001.

[13] M. Ehrig and S. Staab. QOM - quick ontology mapping. In International Semantic Web Conference, pages 683–697, 2004.

[14] F. Baader, D. Calvanese, D. McGuinness, D. Nardi, and P. Patel-Schneider, editors. The Description Logic Handbook: Theory, Implementation and Applications. Cambridge University Press, 2003.

[15] Brewster, C., Alani, H., Dasmahapatra, S., Wilks, Y., 2004. Data driven ontology evaluation. In: Proceedings of International Conference on Language Resources and Evaluation (LREC04), Lisbon, Portugal.

[16] Smith, B., Ceusters, W., Klagges, B., Kohler, J., Kumar, A., Lomax, J., Mungall, C., Neuhaus, F., Rector, A., Rosse, C., 2005. Relations in biomedical ontologies. Genome Biology 6 (R46).