Creating a classification of image types in the medical literature for visual categorization

Henning Müller[ab], Jayashree Kalpathy–Cramer[c], Dina Demner–Fushman[d], Sameer Antani[d]

[a]University of Applied Sciences Western Switzerland (HES–SO), Sierre, Switzerland;
[b]Medical Informatics, University Hospitals & University of Geneva, Switzerland;
[c]Harvard University, Cambridge, MS, USA;
[d]U.S. National Library of Medicine (NLM), NIH, Bethesda, MD, USA

ABSTRACT

Content–based image retrieval (CBIR) from specialized collections has often been proposed for use in such areas as diagnostic aid, clinical decision support, and teaching. The visual retrieval from broad image collections such as teaching files, the medical literature or web images, by contrast, has not yet reached a high maturity level compared to textual information retrieval. Visual image classification into a relatively small number of classes (20–100) on the other hand, has shown to deliver good results in several benchmarks. It is, however, currently underused as a basic technology for retrieval tasks, for example, to limit the search space. Most classification schemes for medical images are focused on specific areas and consider mainly the medical image types (modalities), imaged anatomy, and view, and merge them into a single descriptor or classification hierarchy. Furthermore, they often ignore other important image types such as biological images, statistical figures, flowcharts, and diagrams that frequently occur in the biomedical literature. Most of the current classifications have also been created for radiology images, which are not the only types to be taken into account.

With Open Access becoming increasingly widespread particularly in medicine, images from the biomedical literature are more easily available for use. Visual information from these images and knowledge that an image is of a specific type or medical modality could enrich retrieval. This enrichment is hampered by the lack of a commonly agreed image classification scheme.

This paper presents a hierarchy for classification of biomedical illustrations with the goal of using it for visual classification and thus as a basis for retrieval. The proposed hierarchy is based on relevant parts of existing terminologies, such as the IRMA–code (Image Retrieval in Medical Applications), ad hoc classifications and hierarchies used in imageCLEF (Image retrieval task at the Cross–Language Evaluation Forum) and NLM's (National Library of Medicine) OpenI. Furtheron, mappings to NLM's MeSH®(Medical Subject Headings), RSNA's RadLex (Radiological Society of North America, Radiology Lexicon), and the IRMA code are also attempted for relevant image types. Advantages derived from such hierarchical classification for medical image retrieval are being evaluated through benchmarks such as imageCLEF, and R&D systems such as NLM's OpenI. The goal is to extend this hierarchy progressively and (through adding image types occurring in the biomedical literature) to have a terminology for visual image classification based on image types distinguishable by visual means and occurring in the medical open access literature.

Keywords: medical image classification, information retrieval, image categorization, document image types, image type terminology

1. INTRODUCTION

Content–based image retrieval (CBIR) has received much research attention over the past 20 years,[1,2] including in the medical domain.[3–6] Visual access to the medical literature has also been proposed several times[7,8] but despite many promises, the results in benchmarks have shown a rather low performance of visual retrieval and have always been much better for text retrieval.[9,10] Image classification on the other hand, has consistently delivered fairly good results even for large numbers of classes and hierarchical classification.[11,12]

Further author information: (Send correspondence to Henning Müller, Email: henning.mueller@hevs.ch)

Medical Imaging 2012: Advanced PACS-based Imaging Informatics and Therapeutic Applications,
edited by William W. Boonn, Brent J. Liu, Proc. of SPIE Vol. 8319, 83190P · © 2012 SPIE
CCC code: 1605-7422/12/$18 · doi: 10.1117/12.911186

In the last couple of years several image search engines have become available such as Goldminer,[13] Yottalook*, mainly addressing radiologists, Springerimages† that allows searching in images of all Springer journals, and NLM's PubMedCentral®‡. However all these large-scale image search engines index only the image captions or the article full text to search for images. On a larger scale and particularly in the biomedical domain, the availability of Open Access journals, such as BioMedCentral, has also made available large amounts of medical images for automatic analysis with limited copyright restrictions. Currently, over 1 million images are available via FTP (File Transfer Protocol) from NLM's PubMedCentral § web page. These journals cover a very large variety of images and go far beyond the image types often covered in specialized image retrieval collections, such as the early ImageCLEF or IRMA collections.[10, 11] Images thus are not restricted to classical imaging modalities such as x–ray, CT (Computed Tomography) or MRI (Magnetic Resonance Imaging) but also cover a large variety of rendered and reconstructed multidimensional images such as protein types, genetics–related images, and many forms of flowcharts and graphics. Not all of these images might be relevant for information retrieval and for determining the content of articles with visual means.

To classify medical terms and concepts, a large variety of large–scale terminologies exists. For example, UMLS¶ (Unified Medical Language System) a meta–thesaurus containing several terminologies, such as MeSH‖ (Medical Subject Headings); or SNOMED** (Systemized NOmenclature in MEDicine). Besides these very large terminologies covering most medical domains, there are many specialized terminologies and ontologies for small domains. For radiology, this includes, for example, the ACR code†† (American College of Radiology) created for medical teaching files. A radiology terminology with a broader coverage is RadLex (Radiology Lexicon,[14]) that is currently being developed and extended very actively. Another terminology specific for medical images is the IRMA code[15] that includes modality, information about the views, anatomic region and biological system depicted in the image. The IRMA code has been used for medical image classification in the past. Besides these, there are a few ad hoc terminologies or classifications that have been used for image classification, such as those of the ImageCLEF benchmark in 2011 containing 18 image types. This terminology was largely criticized for being incomplete and ambiguous.[16] These criticisms and reflections among several participants of the image classification challenge have led to the work presented in this article. Whereas SNOMED and UMLS are both copyrighted and not always available free of charge or without restrictions, the terminologies that we map the proposed system to, are all available free of charge: MeSH, the IRMA code and RadLex.

When concentrating on radiology journals alone, an automatic visual classification into radiology modalities seems sufficient and using the IRMA code (or a subset of it as most items are never being used and are potentially hard to distinguish even for trained experts) can be quite useful. Radiology modalities are also well covered in common terminologies such as MeSH or RadLex. When looking at the general biomedical literature, the variety of image types becomes much larger and, to the best of our knowledge, no common classification or terminology for biomedical document images exists. As captions are not always describing the image types well, visual classification could allow enriching the retrieval of images automatically by extracting the required categories. This could lead to better retrieval of targeted information from the medical literature and restrict search or analysis to specific types of images. For example, restricting the search results to only radiology images or simply excluding all graphs or clinically non–relevant images. Mapping the proposed categories onto existing terminologies such as RadLex, MeSH and IRMA can then also help linking images with textual information of semantic categories of ontologies that would allow semantic reasoning based on the data. Similar ideas are behind the LinkedLifeData‡‡ initiative that tries combining and making interoperable many medical terminologies to link data coming from various sources and potentially even from different media. In the context of the semantic

*http://www.yottalook.com/

†http://www.springerimages.com/

‡http://pubmedcentral.gov/

§http://www.ncbi.nlm.nih.gov/pmc/tools/ftp/

¶http://www.nlm.nih.gov/research/umls/

‖http://www.nlm.nih.gov/mesh/

**http://www.ihtsdo.org/snomed-ct/

††http://www.intrad.ch/intrad/acr/

‡‡http://linkedlifedata.com/

web, such semantic interoperability is also an important topic for the domain of biomedical image analysis and retrieval.

This article is structured as follows: Section 2 explains the basic methodology used when writing this text, Section 3 describes the classification terminology itself including an analysis of its current status and ways to extend it. The article finishes with a conclusion in Section 4.

2. METHODS

This article presents a unique system for classification of images in the medical literature and describes a hierarchy of image types occurring in the medical publications. The goal is a hierarchy of visually distinguishable and meaningful figure types. The system is not intended to be exhaustive. Our goal was to develop a small number of classes that will allow visual systems to learn the characteristics of these categories.

The related literature of classifications and terminologies for medical images such as DICOM, RadLex, the IRMA code, the image classes of the ImageCLEF benchmark but also non–radiological systems such as MeSH and UMLS were analyzed to map types of images into a hierarchy. The data occurring in the PubMedCentral Open Access image subset with over one million images from the biomedical literature was taken to validate and extend the hierarchy of classes by classifying 2000 images into the categories manually and noting down all types of occurring problems. The classification also includes ideas germinated from methods developed for combining medical domain ontological knowledge and low–level Image features.[17]

Based on the hierarchy, a set of classes was determined and then mapped onto three medical terminologies that are unconditionally available free of charge: MeSH, RadLex and the IRMA code. These mappings will allow for semantic interoperability as well as reuse of data extracted with visual means for textual information retrieval. All terms are defined in an as detailed manner as possible, basing the definitions, where available, on the MeSH definitions for the related terms. Our main goal was also to base the hierarchy on images existing in the Open Access database. New classes were added when the new images selected for the system development did not fit well any of the existing classes.

3. RESULTS

Based on the available text, a hierarchy of terms or visual classes has been developed, starting with separating images relevant for a clinical analysis, such as medical modalities, and images not directly relevant and rather used to visualize data such as graphs. A category apart, are compound images that contain several sub–figures that should in fact best be analyzed separately, sub–figure by sub–figure. In a first step, it could be sufficient to classify these as compound images. Later on, the set of sub–figure could maybe be classified to the closest category in the hierarchy. The following hierarchy has been created based on the available sources and the images occurring in the medical literature.

3.1. The proposed hierarchy/terminology

To allow for interoperability with other main terminologies, the following codes were added when a correspondence to the closest term in MeSH, RadLex or IRMA code could easily be established.

The MeSH code is added in square brackets *[]*, the IRMA code in curly brackets {}and RadLex in parentheses*()*. For the IRMA code, only the technical axis "T" of the four axes of the code is used; if more details can be given, "x indicates the possible extensions. The hierarchical terms intended to be used for the classification (leaf end notes) are in bold font.

- **Compound/multi–panel images of any type, 2x2, 3x3, ...** [n/a] {T9xxx} (n/a)
- Diagnostic imaging [E01.370.350] {n/a} (RID35725)
  - Radiology images [H02.403.740] {n/a} (RID10345) and nuclear medicine [H02.403.740.500] {n/a} (RID10330)
    - **Ultrasound/echo** [E01.370.350.850] {T2xxx} (RID10326)

- * **Magnetic Resonance Imaging, MRI** [E01.370.350.500] {T3xxx}(RID10312)
  * **Computerized Tomography, CT** [E01.370.350.710.800] {T14xx} (RID10321)
  * **2D Radiography, film or digital** [E01.370.350.700] {T11xx} (RID10349 digital, RID10353 film)
  * **Angiography, radiography of vessels with a contrast agent** [E01.370.350.700.060] {T13xx} (RID10371)
  * **Positron Emission Tomography, PET** [E01.370.350.350.800.700] {T44xx} (RID10337)
  * **Single Photon Emission Computed Tomography, SPECT** [E01.370.350.350.800.800] {T43x} (RID10334)
  * **Combined modalities in one image such PET/CT, PET/MRI, dual energy CT, fMRI** [n/a] {PET/CT T441x, PET/MRI T442x}(RID10341 PET/CT, RID10342 PET/MRI)
  * **Infrared** [E01.370.350.800] {n/a, maybe T56x} (n/a)
- Visible light photography, gross level [E05.712] {T53x} (n/a)
  * Gross photography of organs, tissue [E01.370.350.600] {T53x} (n/a)
    - **Skin** [A17.815]/dermatology [H02.403.225] {T53x} (n/a)
    - **Other organs** [n/a] {T53x} (n/a)
  * **Endoscopy pictures** [E01.370.388.250] {T51xx} (n/a)
- Printed signals, waves [Electrodiagnosis E01.370.405] {T61} (n/a)
  * **EEG** [E01.370.405.245] {T61} (n/a)
  * **ECG / EKG** [E01.370.405.240] {T61} (n/a)
  * **EMG** [E01.370.405.255] {T61} (n/a)
- Microscopic images [E01.370.350.515] {T52xx} (n/a)
  * Light microscopy [UMLS C0430389] {T52xx}(n/a)
  * **Electron microscope** [E01.370.350.515.402] {T524x} (n/a)
    - **Transmission microscope** [E01.370.350.515.402.580] {T524x} (n/a)
  * **Fluorescence images** [E01.370.350.515.458] {T523x} (n/a)
  * **Microscopy, Interference** [E01.370.350.515.513] {T52xx}(n/a)
    - **Phase contrast** [E01.370.350.515.513.569] {T52xx}(n/a)
  * **Dark field** [UMLS C2827982] {T52xx} (n/a)
- Reconstructed/rendered images [n/a] {T744x} (n/a)
  * **3D reconstructions or 3D views** [L01.224.308.410] {T744x}(RID11227)
  * **2D reconstructions** [Image Enhancement L01.224.308.380] {n/a} (n/a)

- Conventional biomedical illustrations [J01.897.280.500.480] {T9xxx} (n/a)
  - Graphs [V01.200] {T9xxx} (n/a)
    * **Tables, forms** [V02.978] {n/a} {n/a} (n/a)
    * **Program listing** [n/a] {n/a} (n/a)
    * **Statistical figures, graphs, pie charts, histograms, other charts ...** [V02.956] {T95xx} (n/a)
    * **Screenshots** [n/a] {T94xx} (n/a)
    * **Flow charts** [L01.224.900.820] {T96xx} (n/a)
    * **System overviews or overviews of components including links and graphics for the parts** [n/a] {T96xx} (n/a)
    * **Gene sequence** [G02.111.570.080] {n/a} (n/a)
    * **Chromatography, Gel** [E05.196.181.400.250] {n/a} (n/a)

    * **Chemical structure** [n/a] {n/a} (n/a)
    * **Symbol** [UMLS:C0679214] {n/a} (n/a)
    * **Mathematics, formulae** [UMLS:C0456604] {n/a} (n/a)
  - **Non clinical photos** [E05.712] {n/a} (n/a)
  - **Hand–drawn sketches** [V01.185.625] {T93xx} (n/a)

It can be seen that the clinical modalities are well represented in all terminologies. The IRMA code that is developed specifically for images is the most complete for our image types, but it also has some classes missing. In several cases, the IRMA classification was ambiguous and the codes could be attributed to several classes in the hierarchy. Note, that most of the classes in our system could be found in the larger systems, as exemplified by the UMLS concept identifiers for the terms that could not be mapped to MeSH.

3.2. Term definitions and examples

These term definitions are partially based on the MeSH or UMLS definitions. Non–existing definitions have been created by us based on the found examples. The goal was to allow for a clear separation between categories to avoid ambiguities wherever possible.

Compound/multi–panel images of any type, 2x2, 3x3, ... An exclusive category that identifies all compound figures as all figures containing more than a single sub–figure. These figures could be separated to be classified into the categories of the sub–figure types in a next step.

Ultrasound/echo The visualization of deep structures of the body by recording the reflections of echoes of pulses of ultrasonic waves directed into the tissues. Use of ultrasound for imaging or diagnostic purposes employs frequencies ranging from 1.6 to 10 megahertz.

Magnetic Resonance Imaging, MRI Non–invasive method of demonstrating internal anatomy based on the principle that atomic nuclei in a strong magnetic field absorb pulses of radiofrequency energy and emit them as radiowaves which can be reconstructed into computerized images. The concept includes proton spin tomographic techniques.

Computerized Tomography, CT Tomography using radioactive emissions from injected Radionuclides and computer Algorithms to reconstruct an image.

2D Radiography, film or digital Examination of any part of the body for diagnostic purposes by means of x–rays or gamma rays, recording the image on a sensitized surface (such as photographic film).

Angiography Radiography of vessels with a contrast agent

Positron Emission Tomography, PET An imaging technique using compounds labelled with short–lived positron–emitting radionuclides (such as carbon–11, nitrogen–13, oxygen–15 and fluorine–18) to measure cell metabolism. It has been useful in study of soft tissues such as cancer; the cardiovascular system and the brain.

Single Photon Emission Computed Tomography, SPECT Single–Photon Emission–Computed Tomography is closely related to positron emission tomography, but uses isotopes with longer half–lives and resolution is lower. A method of computed tomography that uses radionuclides which emit a single photon of a given energy. The camera is rotated 180 or 360 degrees around the patient to capture images at multiple positions along the arc. The computer is then used to reconstruct the transaxial, sagittal, and coronal images from the 3–dimensional distribution of radionuclides in the organ. The advantages of SPECT are that it can be used to observe biochemical and physiological processes as well as size and volume of the organ. The disadvantage is that, unlike positron–emission tomography where the positron–electron annihilation results in the emission of 2 photons at 180 degrees from each other, SPECT requires physical collimation to line up the photons, which results in the loss of many available photons and hence degrades the image.

Combined modalities in one image such PET/CT, PET/MRI, dual energy CT, fMRI Several modalities can be combined or taken in short intervals to show both anatomical and metabolic sides of information on a subject.

Infrared Thermography: Measurement of the regional temperature of the body or an organ by infrared sensing devices, based on self–emanating infrared radiation.

Skin photography A noninvasive diagnostic technique that enables an experienced clinician to perform direct microscopic examination of the surface and architecture of pigmented SKIN lesions.

Photography of other organs Method of making images on a sensitized surface by exposure to light or other radiant energy.

Endoscopy pictures Procedures of applying endoscopes for disease diagnosis and treatment. Endoscopy involves passing an optical instrument through a small incision in the skin i.e., percutaneous; or through a natural orifice and along natural body pathways such as the digestive tract; and/or through an incision in the wall of a tubular structure or organ, i. e. transluminal, to examine or perform surgery on the interior parts of the body.

EEG Electroencephalography; recording of electric currents developed in the brain by means of electrodes applied to the scalp, to the surface of the brain, or placed within the substance of the brain.

ECG / EKG Recording of the moment–to–moment electromotive forces of the heart as projected onto various sites on the body's surface, delineated as a scalar function of time. The recording is monitored by a tracing on slow moving chart paper or by observing it on a cardioscope.

EMG Measurement of muscular activity via electrodes, often shown on screen or on paper and for several electrode placements.

Light Microscopy Using a beam of visible light and optical magnification to image a specimen.This category includes most histopathology microscopy images.

Electron microscope Microscopy using an electron beam, instead of light, to visualize the sample, thereby allowing much greater magnification.

Transmission microscope Electron microscopy in which the electrons or their reaction products that pass down through the specimen are imaged below the plane of the specimen.

Fluorescence images Microscopy of specimens stained with fluorescent dye (usually fluorescein isothiocyanate) or of naturally fluorescent materials, which emit light when exposed to ultraviolet or blue light. Immunofluorescence microscopy utilizes antibodies that are labeled with fluorescent dye.

Microscopy, Interference The science and application of a double–beam transmission interference microscope in which the illuminating light beam is split into two paths. One beam passes through the specimen while the other beam reflects off a reference mirror before joining and interfering with the other. The observed optical path difference between the two beams can be measured and used to discriminate minute differences in thickness and refraction of non–stained transparent specimens, such as living cells in culture.

Phase contrast A form of interference microscopy in which variations of the refracting index in the object are converted into variations of intensity in the image. This is achieved by the action of a phase plate.

Dark field microscopy A microscopic technique in which the light that does not come into contact with the structure or details of interest is subtracted from the ocular image. This yields an image in which the structure or details are alight while the areas where the structures or details are absent are dark.

3D reconstructions or 3D views The process of generating three–dimensional images by electronic, photographic, or other methods. For example, three–dimensional images can be generated by assembling multiple tomographic images with the aid of a computer.

2D reconstructions Improvement of the quality of a picture by various techniques, including computer processing, digital filtering, etc.

Tables, forms Tables with columns of data entries that are the main item of a figure; can be in various formats with horizontal and vertical lines.

Program listing Presentations of nonstatistical data of program code in tabular form.

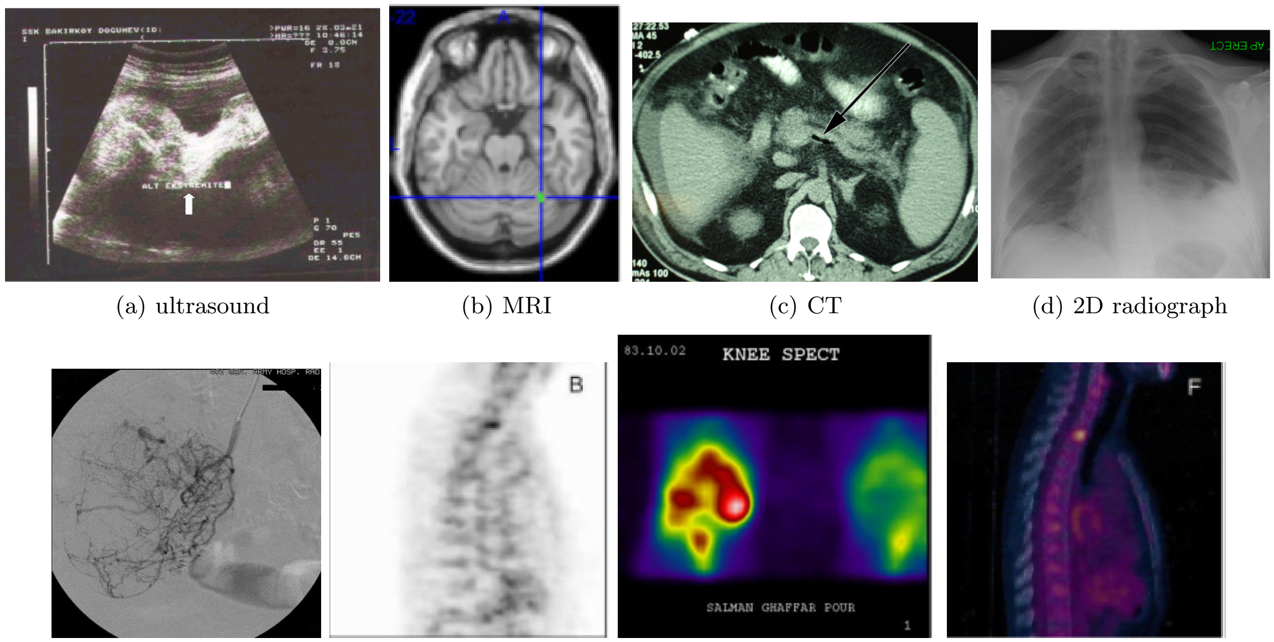
Statistical figures, graphs, pie charts , line charts, histograms, other charts ... Works consisting of presentations of numerical data on particular subjects in a graphical manner in various ways.

Screenshots Screenshot of a software tool showing the graphical user interface of the tool or an output on the user interface.

Flow charts A diagram that shows step-by-step progression through a procedure or system especially using connecting lines and a set of conventional symbols.

System overviews or overviews of components including links and graphics for the parts A global system overview with several distinct parts being part of an overall system; components can be represented with graphics; not a flow chart that has a flow of information or of actions as the main topic.

Gene sequence A sequence of genes as the main part of a figure.



(a) ultrasound (b) MRI (c) CT (d) 2D radiograph

(e) angiography (f) PET (g) SPECT (h) PET/CT

Figure 1. Example images of the radiology category

Chromatography, Gel Column chromatography that separates compounds on the basis of molecular size by using a molecular sieve, usually a zeolite; molecules larger than the largest pore size are excluded from the column.

Chemical structure Figure containing as main object a chemical formula or modeling of several chemical formulae.

Symbol Abstract symbols that are not directly part of a mathematical formula or chemical structure.

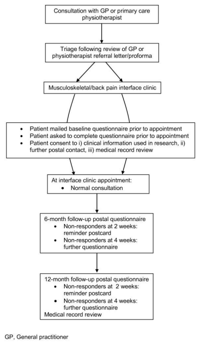
Mathematics, formulae Formulae representing mathematics as main topic of a graphic so not as part of the main text.

Non clinical photos Method of making images on a sensitized surface by exposure to light or other radiant energy; this category takes images not showing lesions, so different from the other photographic categories; this can include images of experimental setups.

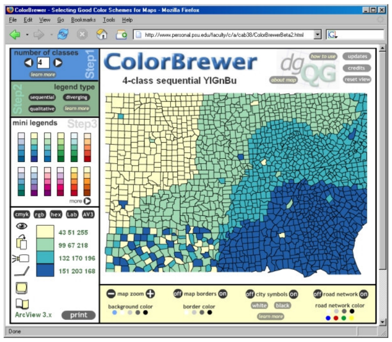
Hand–drawn sketches Drawings, works consisting of graphic representations of objects or ideas by lines; manually written.

We can see in Figure 1 a few example images from the radiology domain. Particularly the classes MRI and CT are often mixed up visually as their characteristics resemble each other.

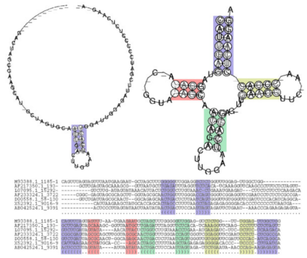
In Figure 2 several examples from the graph (or non–clinical illustrations) category are shown. Particularly the sub–category graphs can have a large variety of figures. Usually many of the graphs are not relevant when representing the content of an article through images. Classifying images into these categories can often help to filter our these non-relevant images, not using them for similarity retrieval. Flow charts or system overviews would require an analysis of the components for a more detailed analysis.



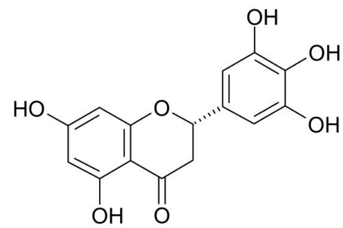
(a) flow chart



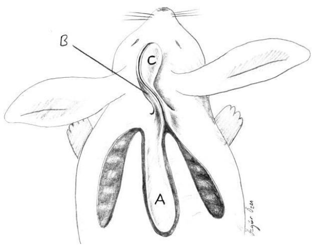
(b) screen shot



(c) gene sequence



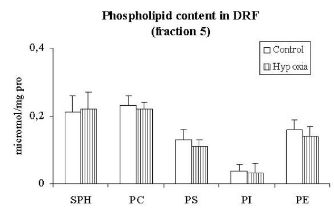
(d) chemical structure



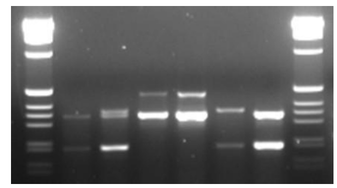
(e) hand drawn sketch

		Event/Disease		
		Yes	*No*	*Total*
Test Result	*Positive*			
	Negative			
	Total			

(f) table

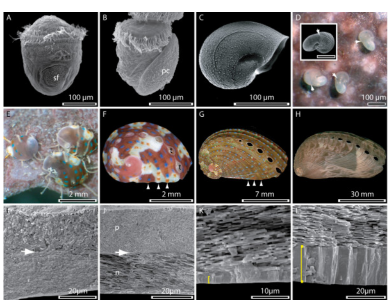


(g) graph

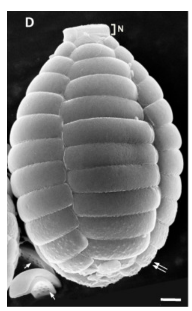


(h) gel

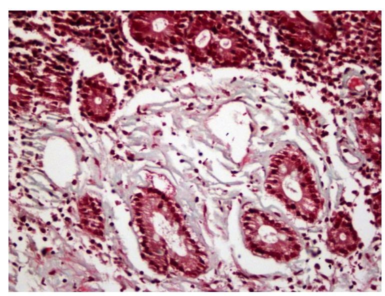
Figure 2. Example images of the very broad graphs category



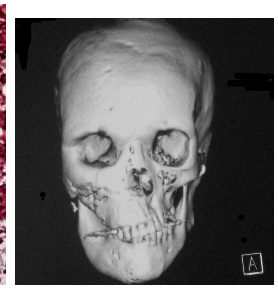
(a) compound figure



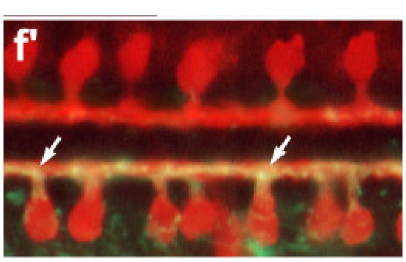
(b) electron microscope



(c) light microscopy

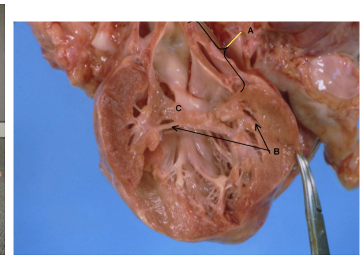


(d) 3D reconstruction



(e) fluorescence microscopy

(f) general photograph



(g) gross pathology photograph

(h) EEG

Figure 3. Example images of various categories mixed

Finally, Figure 3 shows a selection of images from the other categories. Particularly the compound figure category (or multi–panel) contains an extremely large variety of images and will in the future require a more detailed analysis. The different types of microscopy images are important for the biology domain that is often neglected in medical image analysis and retrieval.

3.3. Ground truth creation

The anove hierarchy containing 38 leaf terms has been tested on a subset of 2000 images selected from the PubMedCentral database. The 2000 ground truth images were annotated manually. This data set will be extended and then a semi–automatic classification could increase the size of the available ground truth set, for example, to the over one million images of PubMedCentral. A visual classification accuracy of 90% has been reached in ImageCLEF 2011 with a very similar hierarchy containing 18 classes. Similar results are expected for this set of images. The set will be made available as part of the ImageCLEF 2012 challenge including a defined training set and a test set.

3.4. Advantages and inconveniences of the proposed structure

It is clear that large–scale terminologies such as UMLS, SNOMED and MeSH cover a much broader spectrum of medicine. Many of the categories are also indeed covered already by MeSH and even broader with UMLS, but coding with large terminologies needs more time for learning and the level of detail is often too high, particularly in such fields as visual classification.

Specialized terminologies or hierarchies can thus be much better adapted for a specific problem. This is clearly the case for RadLex with the goal to document radiology cases. This works well in a pure radiology domain, but the biomedical literature is currently vast and many of our defined image types do not occur in the hierarchy of RadLex. For this reason we added several biology–related image types. The IRMA code has a very large coverage of the image types and this is not too surprising as is was created with the goal to categorize medical images into their types along the four axes: imaging technology, view, anatomy and biological system. Despite being specialized, the IRMA code system has many codes and most of them will never get used in clinical practice. In a very large database of over 12'000 images used in the ImageCLEF competition, the number of actually used codes reached 180 of the several thousand (or even million) possible codes and combinations. For the technical or image creation category relevant for this article the number of types was even much lower. In some cases, even an expert could not choose between two codes, which indicates the system is not fully adaptable for the visual classification we aim at with the created hierarchy. In any case, no code can be expected to work in all domains and in practical work it needs to be shown which classifiers are required and where coding difficulties are occurring.

3.5. Extending the existing structure

The current structure was created keeping in mind the idea to add more classes and new categories once this seems necessary and once images appearing in the biomedical literature would require such changes. Visual classification of images is evolving strongly and techniques get increasingly better and can now deal with much larger databases. To create realistic challenges for researchers such a code basis needs to evolve. A first evolution could be the category of compound images that could be used to classify all sub images into their respective classes or already detect the type of compound images (2x2, 3x3, 3x5, etc.). The current set of 38 codes already provides meaningful classes based on the biomedical literature and could scale to potentially millions of images that could be used for information retrieval in a better way.

4. CONCLUSIONS AND FUTURE WORK

This document describes a hierarchy of image types in the biomedical literature. The main goal of the hierarchy is the use for classifying biomedical document images into a defined set of classes that could be useful for human information needs and that can be distinguished visually. So far, textual retrieval has clearly had a better performance compared to visual retrieval but enriching text retrieval with visual classification such as modality information has been shown to improve retrieval quality in the past. When indexing the very varied literature, as it is the case in the biomedical field, specific tasks such as pre–filtering can limit search towards relevant images and allow more options in search interfaces to complement free text search and organize the results set. Mapping this hierarchy to standard terminologies such as MeSH, IRMA or RadLex can equally help to extract low level semantics from the images to improve search systems and move from keyword search to semantic search.

It is clear that such a hierarchy will need to be updated regularly as new image types may appear or the existing image types may change based on new machines becoming available. This hierarchy can also be modified for practical reasons: in case visual analysis is required for a specific application or a specific set of images. Particularly, the creation of training data seems extremely important in such a domain to reach an impact and create a sustainable system. Currently, a training and a test data set are being prepared for the ImageCLEF 2012 image retrieval competition. Thanks to the open access medical literature, very large databases of millions of images are now available for research and extracting information from this visual source that can help information retrieval is one of the goals of this article.

5. ACKNOWLEDGEMENTS

We would like to thank the European Union for their funding in the context of the Khresmoi (grant 257528) and Promise (grant 258191) projects and the Swiss National Science foundation (grant number 205321–130046). This research was also supported by the Intramural Research Program of the Lister Hill National Center for Biomedical Communications, an R&D division of the U.S. National Library of Medicine part of the National Institutes of Health. Jayashree Kalpathy–Cramer was supported in part by a National Library of Medicine grant 5K99 LM009889.

REFERENCES

1. A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content–based image retrieval at the end of the early years," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**, pp. 1349–1380, December 2000.
2. R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Computing Surveys* **40**, pp. 1–60, April 2008.
3. H. Müller, N. Michoux, D. Bandon, and A. Geissbuhler, "A review of content–based image retrieval systems in medicine–clinical benefits and future directions," *International Journal of Medical Informatics* **73**(1), pp. 1–23, 2004.
4. H. D. Tagare, C. Jaffe, and J. Duncan, "Medical image databases: A content–based retrieval approach," *Journal of the American Medical Informatics Association* **4**(3), pp. 184–198, 1997.
5. L. R. Long, S. Antani, T. M. Deserno, and G. R. Thoma, "Content-based image retrieval in medicine: Retrospective assessment, state of the art, and future directions," *International Journal of Healthcare Information Systems and Informatics* **4**, pp. 1–16, January 2009.
6. H. J. Lowe, I. Antipov, W. Hersh, C. A. Smith, and M. Mailhot, "Automated semantic indexing of imaging reports to support retrieval of medical images in the multimedia electronic medical record," *MIM* **38**, pp. 303–307, 1999.
7. T. M. Deserno, S. Antani, and L. R. Long, "Content–based image retrieval for scientific literature access," *Methods of Information In Medicine* **48**, pp. 371–380, July 2009.
8. I. Eggel and H. Müller, "Indexing the medical open access literature for textual and content–based visual retrieval," in *MEDINFO 2010 — Proceedings of the 13th World Congress on Medical Informatics*, **160**, pp. 1277–1281, IOS Press, September 2010.
9. H. Müller, J. Kalpathy-Cramer, C. E. K. Jr., W. Hatt, S. Bedrick, and W. Hersh, "Overview of the ImageCLEFmed 2008 medical image retrieval task," in *Evaluating Systems for Multilingual and Multimodal Information Access – 9th Workshop of the Cross-Language Evaluation Forum*, C. Peters, D. Giampiccolo, N. Ferro, V. Petras, J. Gonzalo, A. Peñas, T. Deselaers, T. Mandl, G. Jones, and M. Kurimo, eds., *Lecture Notes in Computer Science (LNCS)* **5706**, pp. 500–510, (Aarhus, Denmark), September 2009.
10. H. Müller, J. Kalpathy-Cramer, I. Eggel, S. Bedrick, R. Said, B. Bakke, C. E. K. Jr., and W. Hersh, "Overview of the CLEF 2010 medical image retrieval track," in *Working Notes of CLEF 2010 (Cross Language Evaluation Forum)*, September 2010.

11. T. Deselaers, T. M. Deserno, and H. Müller, "Automatic medical image annotation in ImageCLEF 2007: Overview, results, and discussion," *Pattern Recognition Letters* **29**(15), pp. 1988–1995, 2008.
12. H. Müller, T. Deselaers, T. Lehmann, P. Clough, E. Kim, and W. Hersh, "Overview of the ImageCLEFmed 2006 medical retrieval and medical annotation tasks," in *CLEF 2006 Proceedings, Lecture Notes in Computer Science (LNCS)* **4730**, pp. 595–608, Springer, (Alicante, Spain), 2007.
13. C. E. K. Jr. and C. Thao, "Goldminer: A radiology image search engine," *American Journal of Roentgenology* **188**, pp. 1475–1478, 2008.
14. C. P. Lanlotz, "Radlex: A new method for indexing online educational materials," *Radiographics* **26**, pp. 1595–1597, 2006.
15. T. M. Lehmann, H. Schubert, D. Keysers, M. Kohnen, and B. B. Wein, "The IRMA code for unique classification of medical images," in *Medical Imaging 2003: PACS and Integrated Medical Information Systems: Design and Evaluation.*, H. K. Huang and O. M. Ratib, eds., *SPIEProc* **5033**, pp. 440–451, (San Diego, California, USA), May 2003.
16. J. Kalpathy-Cramer, H. Müller, S. Bedrick, I. Eggel, A. S. de Herrera, and T. Tsikrika, "The CLEF 2011 medical image retrieval and classification tasks," in *Working Notes of CLEF 2011 (Cross Language Evaluation Forum)*, September 2011.
17. D. Demner-Fushman, S. K. Antani, M. Simpson, and G. R. Thoma, "Combining medical domain ontological knowledge and low-level image features for multimedia indexing," in *Proceedings of 2nd International Language Resources for Content–Based Image Retrieval Workshop (OntoImage 2008), part of the 6th Language Resources and Evaluation Conference (LREC 2008)*, (Marrakech, Morrocco), 2008.