

# Retrieval from and Understanding of Large-Scale Multi-modal Medical Datasets: A Review

Henning Müller, *Member, IEEE*, Devrim Unay, *Member, IEEE*,

**Abstract**—Content-based multimedia retrieval has been an active research domain since the mid 1990s. In the medical domain visual retrieval started later and has mostly remained a research instrument and less a clinical tool, even though a few tools for retrieval are employed in clinical work. The limited size of data sets due to privacy constraints is often mentioned as a reason for these limitations. Nevertheless, much work has been done in medical visual information retrieval, including the availability of increasingly large data sets and scientific challenges. Annotated data sets and clinical data for the images have now become available and can be combined for multi-modal retrieval. Much has been learned on user behavior and application scenarios. This text is motivated by the advances in medical image analysis and the availability of more public data large data sets that often include clinical data that can be combined for multimodal retrieval based on the experience available in the multimedia community.

This text is a systematic review of recent work (concentrating on the period between 2011-2017) on content-based multi-modal retrieval and image understanding in the medical domain, where image understanding includes techniques such as detection, localization, and classification for leveraging visual content. The main conferences in the field are screened for relevant articles and these are presented in a structured way, identifying current limitations and areas where work is still much required. Objective of the work is to summarize the current state of research for multimedia researchers not working in the medical field. It provides ways to get data sets and identify promising research directions.

The text highlights the areas of advances in the past six years and particularly a trend to use larger scale training data sets as well as deep learning approaches that can replace or complement hand-crafted feature extraction. Using images alone will likely only work in limited sub domains but combining multiple sources of data for multi-modal retrieval has the biggest chances of success, particularly for clinical impact. Future fields of research are identified in the text, as there is a high research potential in the medical multimedia domain.

**Index Terms**—content-based image retrieval, medical images, large scale datasets, multi-modality, big data, deep learning

## I. INTRODUCTION

**T**he medical domain is one of the biggest producers of data. In [1] it is estimated that 30% of world storage was occupied by medical images in 2011, showing the extremely large and often underestimated amount of data produced in medical institutions. Currently, these images are mainly used for the treatment of a single patient and only in limited form across groups of patients [2] or for teaching [3]. Such

secondary use of the data [2] can have many benefits, and content extracted directly from the images is complementary to structured clinical data and free text that are more commonly used in medical decision support.

Content-based medical multimedia retrieval has been active for over 20 years and several review articles have summarized the contributions over the years. In [4], the first systems for visual similarity retrieval are being presented with a relatively small number of references and used techniques. [5] describes the status of research in 2004 with a strong evolution and hundreds of examples of research prototypes and application domains. A more recent overview is [6], published in 2011 and giving an update on new approaches and shortcomings. To our knowledge the latest review article in the field is [7] that was published in 2013 but mainly collects the advances until 2011 and gives a very good overview. Still, approaches like deep learning were not common at the time, even though a few approaches using neural networks have always existed. The data sets have also been much smaller than in the past few years and multimodal retrieval was even less common. Very few user tests have been done with such systems and even fewer systems are really evaluated in clinical practice. A few examples for such user tests do exist [8], [9] that can show an important benefit in clinical practice. In general, many techniques are similar compared to general content-based retrieval [10], [11]. However, the development has been slower and also large data sets are much harder to obtain and annotate [12].

An important part when reviewing the literature is the precise wording of the techniques to be analyzed (included or excluded). The term *retrieval* is most frequently used when the data set does not contain any specific classes or class labels. This is often associated to information *search* as well, where a user has an information need to be fulfilled with documents being relevant or not to the information need. Such a relevance definition is often user- and situation-dependent, thus it changes over time and between persons. Measures for the evaluation are then usually precision and recall. In contrast, the term *classification* implies that the data set can be grouped into a finite set of classes and most often each element is member of a single class. It is also possible to have elements that can belong to several classes, for example when the presence of objects in images is to be classified. The performance measure is most often the classification accuracy in this case, as it is important to know how often a correct label is attached. Precision and recall or the weighted mean of the two are also commonly used. Both approaches can use similar techniques, from visual features to machine learning approaches, but with

H. Müller is with the Information Systems Institute, HES-SO Valais, Sierre, Switzerland, e-mail: henning.mueller@hevs.ch.

D. Unay is with the Biomedical Engineering Department, Izmir University of Economics, Izmir, Turkey, e-mail: devrim.unay@ieu.edu.tr.

Manuscript received December 5, 2016; revised March 1, 2017.

a different scenario and evaluation. The term *localization* is used when inside an image a specific concept or region is to be localized. *Region-of-interest detection* or simply *detection* are often used in a similar way.

In a medical scenario *Computer Aided Diagnosis* (CADx) and *Computer Aided Detection* (CADE) [13] are used and in this case both can refer to retrieval, classification or detection systems that are integrated into the clinical work flow. The more general term of *clinical decision-support* or *computer-based decision support* can also be used for many of the tools described here. As inclusion criteria we specifically analyze these terms and thus take retrieval in a larger sense.

*Image understanding* is in this case a term that covers techniques for detection, localization and classification, and it integrates with retrieval as well. The idea is that higher level content is extracted from the image to help understand the global content of a medical image or a volume. This can be detecting and localizing specific lesions or classifying texture in areas such as the lung tissue into classes that are meaningful for a clinician. Image retrieval is then an interactive way of image understanding as a clinician can query with a case or an image region and find visually similar cases or images with visually similar regions. The found cases are then used for better image understanding. The place of such tools in the clinical work flow is explained in Figure 2.

#### A. Domain-specific Difficulties

The medical domain is like many other specific domains for multimedia analysis. Most of the basic techniques are very similar but there are several details that can totally change a scenario and the way a system can be approached. In medical multimedia analysis in a clinical scenario, images always have a context and some meta data such as why the image was taken, the anamnesis of the patients and in general a text report describing the findings from the image. Medical images are taken under fairly standardized conditions with very similar fields of view and a limited number of standard protocols. Computed Tomography even has a calibration, so a fixed value for its gray levels, calibrated with respect to density of water and air. On the other hand, anatomical differences between humans (small, tall, thin, wide, children, adults, etc.) can be very important and make a fully automated analysis sometimes hard. Diseases change the results seen in images in several interacting ways, depending on predispositions, and there are complex links between the different phenomena.

Even scanners (CT – Computer Tomography, MRI – Magnetic Resonance Imaging, etc.) of the same type can produce different images of the same patient based on the exact model or producer. Medical images are not acquired like in digital cameras with direct sensors that detect light but image quality is linked to many factors. Images are reconstructed from complex raw data (a sensor rotating around the patient) with the use of noise reduction and several other algorithms that can strongly change the images produced and thus also the way to extract information. Many of the main imaging modalities are tomographic, so producing volumetric data that is most often stored in series of slices. Multi-modality imaging, for

example combining CT and PET (Positron Emission Tomography) is common, as different aspects of an organ can be imaged in this way (function for PET, anatomy for CT). This produces very large amounts of data and usually requires registration/alignment of the volumes, as the patient might have moved between data acquisitions. Other difficulties are that the imaging process takes time, sometimes in the range of several minutes, so movement artifacts are common from a patient moving but also from heart or lung motion that can not easily be suppressed. There are other types of information available that are linked to images, such as reports, physiological signals, videos of ultrasound or Doppler sequences, etc. This can be even more complex in a surgical context where instrument locations or surgery videos are tracked in 3D. All this can be taken into account for multi-modal analysis and retrieval using varied data sources.

The biggest problem with medical data is its availability for data analysis because of privacy constraints. Whereas the use for a single patient is fine, its use in a research scenario always requires to make an ethics request (IRB — Internal Review Board) that can add restrictions on data use or patient selection. In [14], [15], various constraints are described when using data from several hospitals in a study, which is necessary when rare conditions are analyzed that are rare in single hospitals. The privacy constraints mean that data are often not shared, even though this is seen as beneficial [12]. If privacy constraints are overcome, intelligent retrieval systems can result in improved tele-healthcare by allowing medical experts to access diagnosed images of similar cases at distant sites [16].

Scientific challenges to measure progress have emerged in many research fields, from information retrieval in the 1960s [17] to medical image retrieval in the early 2000s [18]. The first challenge of the medical image analysis conference MICCAI (Medical Image Computing and Computer Assisted Interventions) started in 2007 [19]. Since then, a web page for challenges in biomedical image analysis was created<sup>1</sup>. More recent benchmarks like those of the VISCERAL project made available large data sets [20] and volume retrieval of thousands of patients [21]. More on scientific challenges is discussed in Section III-D.

Other aspects of medical image data are the often high dimensionality with for example thousands of tomographic slides for a single patient. The data are usually varied with many protocols existing and many types of laboratory analyses or other clinical data types. Integrating these data can be complex and requires multidisciplinary knowledge. Linked to the ethical questions and the high cost to have manual annotations of medical experts, there are often little training data available, further complicating the analysis.

#### B. Motivations for this article

This article is motivated by the fact that the last more or less systematic review of medical multimedia retrieval reports on articles already six years old and many developments have changed the research domain of large-scale multi-modal

<sup>1</sup><https://grand-challenge.org/>

retrieval. Deep learning has been a development of only the last 3–4 years that has totally changed the techniques most commonly employed in scientific challenges and that obtains very good results in many competitions. In combination with GPUs (Graphical Processing Units) larger training data sets and data augmentation techniques have become possible. Scientific challenges allowed to create large data sets by sharing efforts of annotation. Such challenges allow comparing different techniques based on the same grounds. Many funding organizations such as the American NIH (National Institutes of Health) now require the data of funded research to become available for the community and this will likely change future research, as large-scale research on medical image data becomes available.

By reviewing existing work and missing parts the text should give multimedia researchers several possibilities to work on available medical data and to respond to current research challenges. It is clear that most medical data sets used are not really large scale but data production in medical institutions is extremely is high. Also the multimodal nature of medical data is often not fully exploited, but possibilities are enormous when the full available data are used.

## II. METHODOLOGY

To find candidate articles for this review we concentrated on conferences and journal articles from 2011–2017, so with little overhead compared to previous review articles. A few older articles were also added, as they were still seen as important with a strong influence on the current work, but they are only used in limited form in the technical sections.

For the articles selected in this survey we systematically analyzed conferences in the field of medical image analysis, without being exhaustive, namely:

- MICCAI;
- IPMI (Information Processing in Medical Imaging);
- ISBI (International Symposium in Biomedical Imaging);
- SPIE Medical Imaging.

In terms of scientific journals we concentrated on the following journals but also searched in the PubMed<sup>2</sup> literature database to complete the initial set of papers:

- Medical Image Analysis;
- IEEE Transactions on Medical Imaging;
- Computerized Medical Imaging and Graphics;
- Journal of Digital Imaging;
- ACM Transactions on Multimedia;
- Radiology;
- Magnetic Resonance Imaging;
- Magnetic Resonance in Medicine;
- Neuroimage.

These sources were manually checked for the past six years and extended by other publications that we were aware of in other sources. The publications are scattered around an extremely large number of sources, as this is a very multi-disciplinary domain. The search terms employed were “content-based medical retrieval”, “medical image retrieval”,

“medical visual information retrieval”, “deep learning”, “convolutional learning”, “large scale medical retrieval”, “medical big data retrieval”, and the like.

The choices for a more detailed analysis can only be a starting point for the references of a complete review. In general we tried to have diversity in sources and contents. Very similar articles of the same authors are discarded. In this case journal articles are preferred over conference papers and more recent articles over older texts. The found references are structured into several categories, starting with the data sets used, clinical application scenarios, techniques used and specific aspects related to scalability of the approaches.

## III. DATA AND GROUND TRUTH AVAILABLE

This section starts with the foundation of databases that have been used for retrieval and then extends to the scientific challenges, ground truth generation, crowdsourcing efforts and clinical scenarios in the field.

Categorization of the found references are roughly done by the scale (small: <1000 data items, medium, or large: >10000), and the type of the databases used (uni-modal comprising 2D and 3D images from single or multiple image modalities, and multi-modal combining images with textual information or other data types). Table I presents a summary of the references respecting this categorization, while Fig. 1 shows a scatter plot of the found references by size and type of the database and publication year. The graphic shows that we have more publications in 2015 and 2016 and larger data sets are mainly used in the last two years, with a focus on 2016. There are many datasets and a majority is still unimodal, but increasingly data types are being mixed for research. Still, the size of the used data sets remains relatively small for retrieval, whereas for other applications larger data sets have been employed.

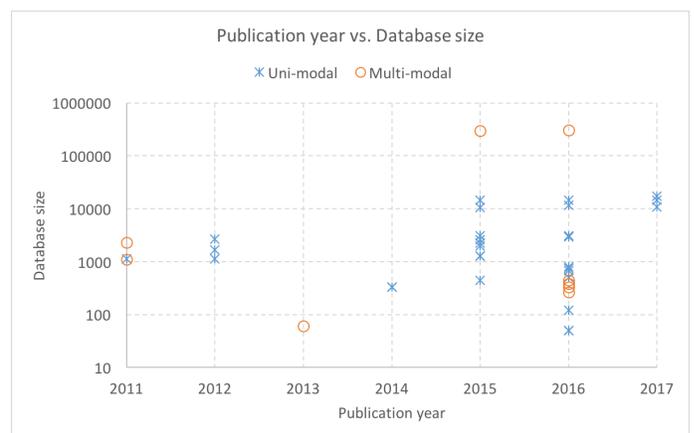


Fig. 1. Scatter plot of database size vs. publication year for medical CBIR.

### A. Small-Scale Uni-Modal/Cue (Traditional) Databases

A variety of medical image repositories have been made available free of charge for research and education purposes.

<sup>2</sup><http://www.pubmed.gov/>

TABLE I  
STUDIES GROUPED BY DATABASE.

| Type of database | Data type              | Scale of database | Studies   |
|------------------|------------------------|-------------------|---|
| Uni-modal        | 2D                     | small             | general jpegs: [22]; microscopy: [23].                                    |
|                  |                        | medium            | mammograms: [24]; x-rays: [25], [26]; microscopy: [27]–[29]; mixed: [30]. |
|                  | 3D                     | large             | mammograms: [31], [32]; x-rays: [33]–[36].                                |
|                  |                        | small             | CT: [37]–[40]; MRI: [41]; PET: [42]; MRI+PET: [43], [44].                 |
| Multi-modal      | 2D + text/demographics | medium            | CT: [45]; CT+PET: [46], [47]; endoscopic video: [48], [49].               |
|                  |                        | large             | microscopy: [50].   |
|                  | 3D + text/topic word   | small             | x-rays: [51].   |
|                  |                        | medium            | mammograms: [52]; retinal images: [52].                                   |
|                  |                        | large             | general jpegs: [53]; mixed: [54].   |
|                  |                        | small             | CT: [55], [56]; MRI: [57]; MRI+PET: [56].                                 |

IRMA (Image Retrieval in Medical Applications)<sup>3</sup>, NCIA (National Cancer Imaging Archive)<sup>4</sup>, ADNI (Alzheimer’s Disease Neuroimaging Initiative)<sup>5</sup>, OASIS (Open Access Series of Imaging Studies)<sup>6</sup>, and ImageCLEF (Multimedia Retrieval in CLEF)<sup>7</sup> are a few examples. ImageCLEF has created at least 20 medical data sets of varying size.

As for retrieval of medical images, several of these repositories have been used by researchers (sometimes only subsets). Recently, [39] employed 50 cases of liver CT images from ImageCLEF for retrieval, while [42] used 331 cases of PET images, [43] and [44] respectively utilized 331 and 805 cases of PET+MR data from ADNI, and [41] MR data of 50 patients from the publicly available PROstate Mr Image SEgmentation (PROMISE) challenge<sup>8</sup>. Using standard databases, even when small, can help to make research reproducible. Unfortunately, sometimes only a subset of the data is being used, which reduces the reproducibility.

Besides these databases several other 2D (Jpeg: [22], microscopy: [23]) and 3D (CT: [37], [38], [40]) image collections, mostly in-house acquired, have been used for medical retrieval. [22] utilized 500 Jpeg’s to realize anatomical categorization from similar cases, while [23] validated retrieval-based breast cancer detection over 120 microscopy images. [37], [38], [40] performed categorization of around 700 chest CT volumes. Besides being small, the data sets can often not be shared and thus research performance can not be judged and is not reproducible.

### B. Medium–Large Scale Uni–Modal/Cue Databases

Several retrieval studies have employed medium– to large–scale subsets of uni–modal/cue image repositories that are available on the web. Among the found references [24], [31], [32] exploited mammogram ROIs for breast cancer detection from the Digital Database for Screening Mammography (DDSM) repository<sup>9</sup>; [26], [33]–[36] utilized X–rays from the IRMA repository; [25] used X–rays from the second National Health and Nutrition Examination Survey (NHANES–II) repository<sup>10</sup> to diagnose vertebrae irregularity; and [45]

employed CTs from the Lung Image Database Consortium (LIDC) collection<sup>11</sup> for pulmonary nodule retrieval.

Other studies mostly used in–house databases at medium–scale. [27], [28] realized breast cancer detection and [29] carried out diagnosis of blood (lymphoma) and nerve tissue (neuroblastoma) cancers from 2D microscopy images, while [46], [47] realized case retrieval from thoracic PET–CT data, [48], [49] executed similar video retrieval of endoscopic videos, and [30] performed fracture categorization using images of multiple modalities (with X–rays dominating). An exception is the work of [50], where a large–scale database of 3D microscopy images (17,107 images) is used for neuron morphology retrieval. The studies show a tendency to have a larger variety of image sources. In previous surveys the focus of medical image retrieval was strongly on radiology, whereas now many other image types have become available digitally.

### C. Medium–Large Scale Multi–Modal/Cue Databases

The plethora of non–imaging data present in various forms (health records including radiology reports, audio, unstructured text, books, journals, taxonomies, ontologies, etc.) bring in valuable information complementary to visual content in images/videos.

Accordingly, several efforts have been made to exploit this complementary information for improved medical retrieval. [51] employed deep learning features extracted from X–rays of 443 cases with demographic information for chest pathology retrieval. [52] used demographics and medical history of patients with image features for diabetic retinopathy diagnosis of 1112 cases from an in-house collected database and for mammography screening of 2277 cases from the DDSM repository. [57] augmented image features from 60 MRIs of ADNI and OASIS with contextual information (demographics, medical history, lab results, and ontologies) for detection of Alzheimer’s and dementia. [55] combined textual information from radiology reports with image features from 265 chest volumes for detection of interstitial lung disease subtypes. [56] used topic words with image features extracted from 379 CT volumes of the I–ELCAP (International Early Lung Cancer Action Program) collection<sup>12</sup> and 331 PET+MRI data of the ADNI collection for detection of lung disease and Alzheimer’s, respectively.

<sup>3</sup><https://ganyomed.imib.rwth-aachen.de/irma/>

<sup>4</sup><https://imaging.nci.nih.gov/ncia/>

<sup>5</sup><http://adni.loni.usc.edu/>

<sup>6</sup><http://www.oasis-brains.org/>

<sup>7</sup><http://www.imageclef.org/>

<sup>8</sup><https://promise12.grand-challenge.org/>

<sup>9</sup><http://marathon.csee.usf.edu/Mammography/Database.html>

<sup>10</sup><https://www.cdc.gov/nchs/nhanes/nhanesii.htm>

<sup>11</sup><https://imaging.cancer.gov/programsandresources/informational/systems/lidc/>

<sup>12</sup><http://www.via.cornell.edu/databases/lungdb.html>

Two large-scale (ImageCLEF) studies complementing image features with textual information extracted from articles have been proposed recently [58]. [53] exploited term suggestions from a medical thesaurus for case-based retrieval in a database of over 300,000 images with associated texts, while [54] benefited from text of articles for modality classification that is also often used to limit the search space for visual retrieval.

#### D. Scientific Challenges

As mentioned beforehand, practically all medical imaging conferences now organize scientific challenges where participants can compare their approaches on a single data set and with a clear and given evaluation scenario. This allows to really compare all the techniques on the same grounds, which is a major change in this field. Such data papers often generate many citations [59] and can thus have a strong impact. One of the first medical multimedia retrieval challenges was ImageCLEF that started with a medical task in 2004 [60]. Since then, many medical and non-medical challenges were held in the context of ImageCLEF. Other non-medical retrieval challenges include TrecVid [61] and MultimediaEval [62].

The Web page on Grand Challenges in Medical Imaging has a list of past and currently active challenges in medical image analysis. Professional platforms such as Kaggle<sup>13</sup> and TopCoder<sup>14</sup> also provide a platform for data providers to share challenges and data with the community and obtain results from many researchers. This is done in a professional way with price money and it does not always include publications of the results.

To just name a few challenges, the MICCAI liver segmentation competition in 2007 had 10 liver volumes and 14 participants [19] that had to segment the data in three hours on site. This might not seem much, but at the time it was more than most other data sets used for the task and it made results really comparable. ImageCLEF has developed multi-modal collections between 600 images in 2004 and 300,000 images in 2013, all with associated text data. A more recent one is the BRATS (Brain Tumor Segmentation) 2014 challenge [63] that allowed comparing algorithms on various parts of tumors. Camelyon16<sup>15</sup> and Tupac (Tumor proliferation and assessment challenge)<sup>16</sup> are two recent challenges on whole slide histopathology images. The 2017 Camelyon challenge distributes over 2 TB of image data, but these are classifications and not directly retrieval challenges.

Basically all major scientific conferences in data analysis now offer challenges and thus a platform to work towards a common task on the same data and compare results. This will likely improve scientific work and foster collaborations.

The VISCERAL project developed challenges based on the concept of moving the algorithms to the data [64] by running challenges in the cloud [65]. This avoids shipping data and allows to test executable code in the exact same environment.

It allows creating a silver corpus by label fusion of results of existing tools that are run on non-annotated data [66]. The concept of Evaluation-as-a-Service (EaaS) [67] is a logical consequence of this and several other medical challenges are now run in a similar way, avoiding physical distribution of data to participants. The Mammography Dream Challenge<sup>17</sup> for example makes 640,000 mammographies of 86,000 patients accessible. The data can be used for analysis but cannot be viewed or downloaded by the participants.

#### E. Ground Truth Generation and Crowd Sourcing

As the manual annotation of images is expensive there are several possibilities to limit the manual or paid effort. The LabelMe<sup>18</sup> game worked on general images and allowed shared annotations of stock photography. While for general objects laypersons can easily be used, for medical data annotation most often trained specialists are needed that makes the problem even harder. If precise tasks are picked, then users can be trained to label also very specific medical images when strict quality control is used [68]. Crowdsourcing [69] has mainly been used for non-medical data annotation, but several projects also on medical data exist with good success [68], [70]. Another option is to use outputs of several systems and then employ label fusion [66]. This is similar to weak labels often used from web data and unsupervised approaches for data enrichment that can work if sufficient amounts of data are available.

#### F. Clinical scenarios and applications

In a clinical scenario, following a medical doctor's order for imaging, the radiologist performs a detailed quality check and analysis of the acquired image and generates a report with the findings (Fig. 2). Medical image retrieval and related techniques can play various roles at different steps of this work flow from image reading to report writing.

Most available multi-modal medical image retrieval systems have so far used large data sets from the medical literature as these images are available in large quantities and can be used relatively easily. Fig. 3 shows an example interface for retrieval of images from the literature developed by the National Library of medicine in the USA. The interface allows keyword search and filtering by image type as well as visual similarity search in the data.

A different interface, shown in Fig. 4, aims at retrieval of similar volumes from tomographic images in radiology [71]. The system is similar to a radiology viewing station and concentrates on a single screen, showing the volume to be diagnosed and a selection of volumes that contain similar regions of interest and the related radiology reports for multi-modal retrieval enriched by semantics.

Table II presents the retrieval publications analyzed between 2011 and 2017. The table shows that there is a large variety of different medical areas that are in the focus of the retrieval.

<sup>13</sup><http://www.kaggle.com/>

<sup>14</sup><http://www.topcoder.com/>

<sup>15</sup><https://camelyon16.grand-challenge.org/>

<sup>16</sup><http://tupac.tue-image.nl/>

<sup>17</sup><https://www.synapse.org/#!/Synapse:syn4224222/wiki/401743>

<sup>18</sup><http://labelme.csail.mit.edu/Release3.0/>

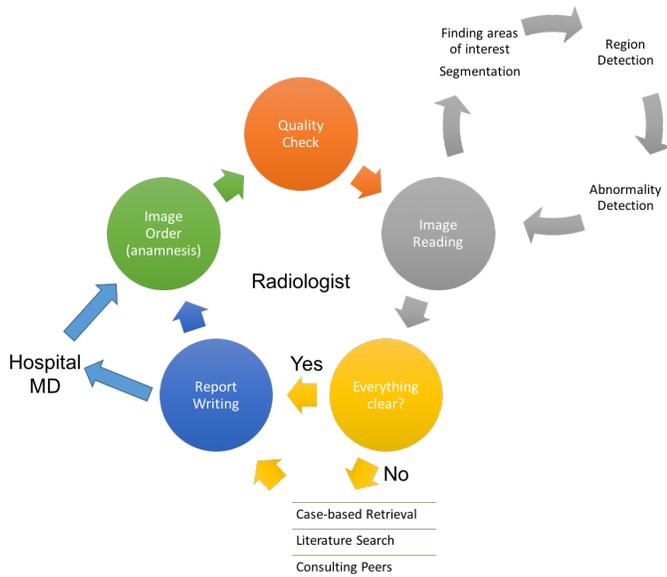


Fig. 2. Example of a radiologist’s work flow in a clinical environment.

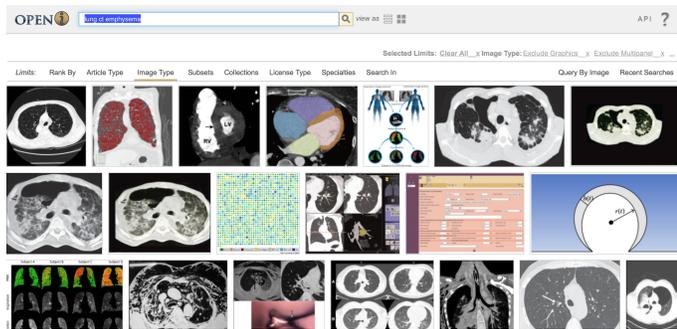


Fig. 3. Screen shot of Open-I to search figures of the biomedical literature.

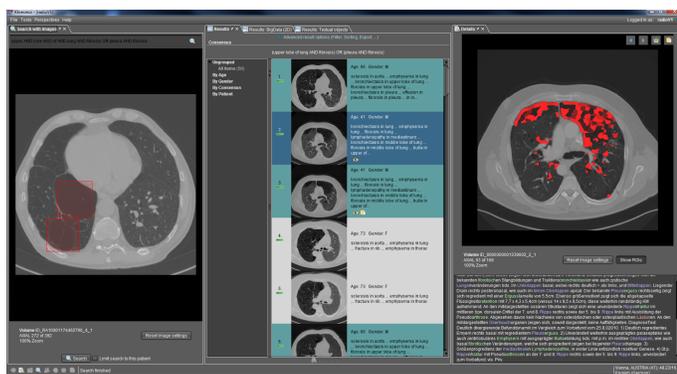


Fig. 4. Screen shot of the Khresmoi radiology search system.

Breast cancer analysis and interstitial lung diseases are two areas where much research has been done in the past and where several data sets are available. This is well represented in the table with several papers. The largest databases are from the biomedical literature and not concentrated on a single medical problem, namely the ImageCLEF databases. ADNI [72] has also made available large data sets, and several

subsets are being used for retrieval in Alzheimer’s patients. In recent years an increasing number of endoscopy videos have been used for retrieval, and also microscopy images such as histopathology have been utilized for retrieval several times [73].

IV. TECHNIQUES USED

This section details our analysis of the found references with respect to the techniques used, particularly by the image/video and non-image features utilized and the machine learning and deep learning approaches employed. The section is completed with a brief presentation on the efforts in industry, as this is also important.

A. Image/Video Features

Similar to [74], we group image/video descriptors into three operational categories: general, mixed, and specialized. Fig. 5 presents the articles based on this categorization.

General descriptors are common in all CBIR systems like color, texture, histograms, shape, etc. These descriptors can be extracted globally over the entire image or locally. General descriptors employed are pixel intensities [22], [26], [29], [30], histograms [24], [25], Haralick’s gray level co-occurrence matrix-based features [24]–[26], [29], wavelets [24], [25], [30], [37], scale invariant feature transform (SIFT) [23], [26]–[28], [30]–[32], [38], GIST [27], [32], histogram of oriented gradients (HOG) [27], local binary patterns (LBP) [33], [36], [38], [75], and Fourier descriptors [25]. Some of these studies represented general descriptors in a bag of visual words model as well [28], [30]–[32], [37].

Mixed descriptors exploit text input or annotations with general descriptors. To augment image information with text, [54] employed text extracted from scientific articles and [55] used text from radiology reports. The multi-modal retrieval system in [53] supports medical information discovery by fusion of general descriptors from images with term suggestions from a medical thesaurus over the text from scientific papers [76]. In order to overcome the subtle visual variations between anatomical structures and obtain better results, [56] used latent semantic topic description with SIFT image features. [51] employed augmented deep learning features with age and gender information of the patients for improved chest pathology retrieval. [57] used shape representation of lateral ventricles with contextual information of demographics, medical history, clinical results and semantic ontologies. [52] employed contextual information like demographics and medical history of the patient with wavelet features extracted from images to improve diabetic retinopathy diagnosis and mammography screening.

Specialized descriptors are defined in [74] as exploiting interrelations between features through incorporation of domain knowledge. We extend this definition and include studies that employ new feature extraction techniques such as features derived from deep learning [34], [35], [38], [41], [42], [51]. Fig. 5 shows that general features are the large majority of image descriptors used for retrieval. Recent years have seen more mixed descriptors and a few based on deep learning but both areas are likely to increase quickly over the coming years.

TABLE II  
MAIN CHARACTERISTICS OF MEDICAL CBIR SYSTEMS

| Ref. | Descriptors | Similarity measure                       | Segmentation | RF  | Modality                 | DB size    | Medical problem                      |
|------|-------------|--|--------------|-----|--------------------------|------------|--------------------------------------|
| [37] | General     | Vector distance                          | -            | No  | CT                       | 675        | Chest disease diagnosis              |
| [38] | Specialized | Vector distance<br>Classifier based      | -            | No  | CT                       | 443        | Chest disease diagnosis              |
| [33] | General     | Classifier based                         | -            | No  | X-ray                    | 14410      | Anatomical categorization            |
| [26] | General     | Vector distance<br>Classifier based      | -            | No  | X-ray                    | 2600       | Anatomical categorization            |
| [24] | General     | Vector distance<br>Perceptual similarity | -            | No  | Mammogram                | 2919 ROIs  | Breast cancer diagnosis              |
| [25] | General     | Vector distance                          | Automatic    | No  | X-ray                    | 2220       | Vertebrae irregularity diagnosis     |
| [40] | General     | Classifier based                         | -            | No  | CT                       | 746 ROIs   | Lung cancer diagnosis                |
| [28] | General     | Vector distance                          | -            | No  | Microscopy               | 3121       | Breast cancer diagnosis              |
| [27] | General     | Vector distance<br>Classifier based      | -            | No  | Microscopy               | 3121       | Breast cancer diagnosis              |
| [42] | Specialized | Classifier based                         | -            | No  | PET                      | 331        | Alzheimer diagnosis                  |
| [22] | General     | Vector distance                          | -            | Yes | Jpeg                     | 500        | Anatomical categorization            |
| [31] | General     | Vector distance                          | -            | No  | Mammogram                | 10553 ROIs | Breast cancer diagnosis              |
| [32] | General     | Vector distance                          | Automatic    | No  | Mammogram                | 11533 ROIs | Breast cancer diagnosis              |
| [55] | Mixed       | Vector distance                          | Automatic    | Yes | CT                       | 265        | Interstitial lung disease diagnosis  |
| [39] | General     | Vector distance<br>Classifier based      | -            | No  | CT                       | 50         | Semantic annotation                  |
| [54] | Mixed       | Classifier based                         | -            | No  | Mixed                    | 306539     | Modality classification              |
| [43] | General     | Vector distance                          | Automatic    | No  | MRI + PET                | 331        | Alzheimer diagnosis                  |
| [44] | General     | Classifier based                         | -            | No  | MRI + PET                | 805        | Alzheimer diagnosis                  |
| [56] | Mixed       | Vector distance                          | Interactive  | No  | CT, MRI + PET            | 379, 331   | Lung cancer, Alzheimer diagnosis     |
| [23] | General     | Graph based                              | -            | No  | Microscopy               | 120        | Breast cancer diagnosis              |
| [48] | General     | Vector distance                          | -            | No  | Endoscopic video         | 1276       | Quick access to similar videos       |
| [49] | General     | Vector distance                          | -            | No  | Endoscopic video         | 1276       | Quick access to similar videos       |
| [45] | General     | Vector distance                          | Interactive  | No  | CT                       | 2003       | Pulmonary nodule retrieval           |
| [53] | Mixed       | Vector distance                          | -            | No  | Jpeg                     | 300000     | Biomedical literature search         |
| [50] | General     | Vector distance                          | -            | No  | Microscopy               | 17107      | Morphological retrieval of neurons   |
| [29] | General     | Classifier based                         | -            | No  | Microscopy               | 1666       | Blood, nerve tissue cancer diagnosis |
| [30] | General     | Vector distance                          | -            | No  | Mixed                    | 2690       | Fracture categorization              |
| [57] | Mixed       | Vector distance                          | Automatic    | Yes | MRI                      | 60         | Alzheimer/Dementia diagnosis         |
| [51] | Specialized | Vector distance<br>Classifier based      | -            | No  | X-ray                    | 443        | Chest pathology retrieval            |
| [34] | Specialized | Vector distance                          | -            | No  | X-ray                    | 14410      | Anatomical categorization            |
| [41] | Specialized | Vector distance                          | -            | No  | MRI                      | 50         | Prostate cancer diagnosis            |
| [35] | Specialized | Vector distance                          | -            | No  | X-ray                    | 10902      | Anatomical categorization            |
| [36] | General     | Vector distance                          | -            | No  | X-ray                    | 14410      | Anatomical categorization            |
| [52] | Mixed       | Classifier based                         | -            | No  | Retinal image, Mammogram | 1112, 2277 | Diabetic retinopathy, Breast cancer  |
| [46] | General     | Classifier based                         | Automatic    | No  | PET + CT                 | 1134       | Lung cancer diagnosis                |
| [47] | General     | Classifier based                         | Automatic    | No  | PET + CT                 | 1134       | Lung cancer diagnosis                |

RF: Relevance feedback. CT: Computed tomography; MRI: Magnetic resonance imaging; PET: Positron emission tomography.

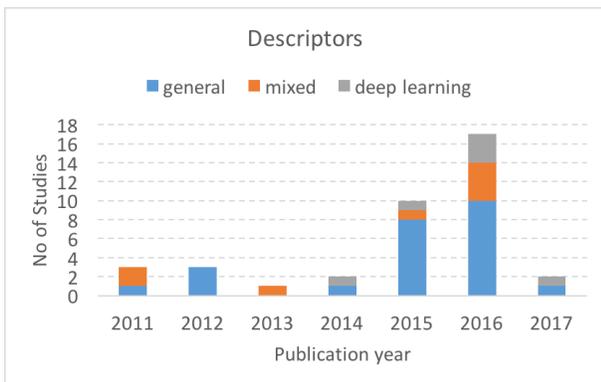


Fig. 5. Descriptors used in the medical CBIR systems.

### B. Non-Image Features

In the scope of multi-modal medical retrieval, studies that exploit non-image features such as text typically make use of natural language processing techniques to mine relevant

information from radiology reports or scientific articles. While a few used standard approaches like term frequency and inverse document frequency (*tf.idf*) analysis [54], [55], others employed more advanced approaches like interactive query expansion using medical thesauri [53]. Some researchers consider images as visual words and applied semantic analysis to evaluate each word's discriminative power [56].

Besides the text based features mentioned above, the references generally employed non-image features that carry information on the patient and the pathology at hand. Typically, these features are expressed in qualitative, semi-quantitative or quantitative scales that need to be converted to numerical values for processing. For example, demographic information such as age, gender, and socio-economic status of subjects is an informative feature often used [51], [52], [57]. Family medical history and clinical features like disease symptoms, examination results, etc. are used as well [52], [57]. In order to improve retrieval performance, [57] benefited from structured representation of anatomical and disease information stored

in ontologies such as the Human Disease ontology<sup>19</sup> and the Foundational Model of Anatomy ontology<sup>20</sup>.

### C. Machine Learning

Machine learning is grouped by the nature of learning signals available as unsupervised (unlabeled data), semi-supervised (partially labeled data), and supervised (fully labeled data). Here, we present the found references that employ machine learning in accordance with this grouping. Fig. 6 shows that a large majority used supervised learning but that a few semi- or unsupervised approaches have started in the past year in the medical field.

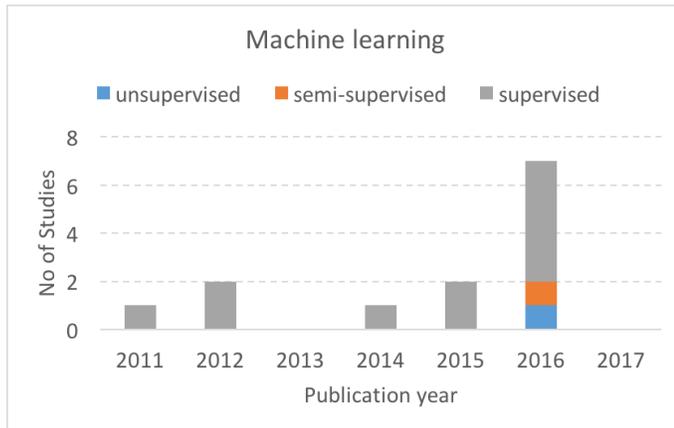


Fig. 6. Machine learning techniques used in the medical CBIR systems.

Among studies employing unsupervised learning, [26] benefited from sparse dictionary learning via the use of  $k$ -singular value decomposition, while [23] used graph-based ranking. Despite most real-world scenarios being semi-supervised, this learning approach is the least frequently used in medical CBIR systems. The only example found is the work of [44] where adaptive ensemble manifold learning is exploited.

On the other hand, supervised learning such as the nearest neighbor algorithm [29], [42], support vector machines [29], [33], [38], [39], [51], [54], discriminative models (differential scatter discriminant criterion) [40] and others like joint kernel based supervised hashing [27] are most commonly used.

### D. Deep Learning

Deep learning has become one of the major hype topics in machine learning [77] and it has had major success in many scientific challenges, from the ImageNet challenges [78] to many other machine learning challenges held. In the medical field it has been employed for detection, segmentation and other decision support tasks [79]–[82]. Our review seems to indicate that for medical image retrieval the use of deep learning has only recently started and is likely to increase strongly over the coming years. In histopathology two reviews highlight the use of deep learning for several tasks [83], [84].

When analyzing the literature we did not find a single article using deep learning directly for medical image retrieval. However several studies made use of deep learning techniques in a retrieval framework for extracting image features. [42] exploited deep learning features for PET retrieval based Alzheimer’s diagnosis, while [38], [51] employed it for pathology retrieval of chest X-rays. [34], [35] used deep learning features for anatomical categorization of X-rays and [41] for prostate cancer diagnosis using MRI retrieval.

Given the extremely high popularity of deep learning in academic research and industry, we expect an influx of literature exploring performance of different deep learning strategies for large-scale retrieval of medical data sets in the coming years.

### E. Efforts of the Industry

The increasing amount and distributed nature of medical data produced, require efficient and intelligent solutions that can transform healthcare [85]. Recognizing this, the industry has started developing and providing dedicated solutions.

IBM Watson Health<sup>21</sup> is a modular system designed around DeepQA, a software architecture for deep content analysis and evidence-based reasoning. It aims at translating medical information into knowledge for more informed decision-making in healthcare by analyzing and interpreting data in various forms (unstructured text, images, audio, and video), providing personalized recommendations, learning from past experiences with the use of machine learning, and allowing interaction via chat bots that engage in dialog.

SAP Medical Research Insights<sup>22</sup> is a web application dedicated to cancer research aiming at easier and more efficient information search for oncologists. The goal is to provide physicians access to every single piece of information (images, patient records, laboratory results, reports, etc., so multimodal data) about cancer patients stored in separated data sources possibly at different physical locations.

Siemens Theseus-Medico<sup>23</sup> was a large-scale national project in Germany across several industries and Siemens led the medical part of the project. The objective was to foster semantic technologies for the extraction and interpretation of knowledge from free text, signals and image data. Many decision support tools for automatic data annotation and report generation were created as prototypes in this project, including semantic retrieval systems.

## V. SCALABILITY ASPECTS OF THE APPROACHES

The analysis of the presented approaches shows that most published articles still use relatively small data sets even though much larger data are becoming available. Moreover, scalability of the presented approaches does not seem a major concern. Among the studies analyzed only a small subset addressed scalability through the use of hash-based algorithms [27], [28], [32] and asymmetric binary coding [50].

<sup>21</sup><http://www.ibm.com/watson/health/>

<sup>22</sup><https://icn.sap.com/projects/sap-medical-research-insights.html>

<sup>23</sup><http://www.digitale-technologien.de/DT/Redaktion/EN/Downloads/Publikation/theseus-forschungsprogramm-broschuere-en.>

<sup>19</sup><http://www.obofoundry.org/ontology/doid.html>

<sup>20</sup><http://si.washington.edu/projects/fma>

Broader research infrastructures may need to become available for data analysis on much larger resources. The genomics domain has shown how such a creation of large-scale computing infrastructures can modify a field. Maybe the combination of genomics with imaging, RadioGenomics [86], can also help the imaging field to address challenges on larger data and of a more multi-modal nature.

Hospitals will require large computing infrastructures for precision medicine and this can also lead to approaches similar to EaaS, where the code is moved towards the data and not the data to the code. This can strongly help with data confidentiality challenges and potentially allow to run scientific challenges fully inside medical institutions.

## VI. DISCUSSION

This review has limitations and the number of analyzed texts from 2011 to 2017 is limited to only six years. The difficulty with the domain is that publications are scattered around many different research fields and thus it is hard to be systematic and complete. Some publications are in small workshops that can easily be missed and are thus not considered in the analysis. Still, a few trends become visible and other trends can be expected for the coming years. This text shows several shortcomings of the medical domain compared to more general content-based multimedia retrieval, such as small data sets and often the limited use of multi-modal data.

### A. Scalability

Even though the databases in medical multimedia analysis have increased in size over the past years there is still an alarmingly strong use of small data sets that leads to problems with variety and interpretability of outcomes [87]. It is important to make sure that publications use standard data of sufficient size in the future, so meaningful results are obtained. This will create a need to work on scalability and efficient algorithms for data analysis, at least for the online phase.

### B. Security and Privacy

As big data transforms healthcare, governance questions such as ownership, privacy, security, and standards emerge [88]–[90]. An appropriate balance between public health and data privacy needs to be ensured [91] to allow exploiting data for the good of society and still protect individual person's privacy. Models that move the algorithms towards the data can help with protection, as only algorithms can access data and not persons. Physically having the data may in fact become unnecessary when data sets become extremely large. Cloud computing (public or private) may in this case be a driver behind research in medical visual information retrieval.

### C. Clinical Benefits

We are in the age of big data which is characterized by big volume, large data variety, high velocity of data generation/update and high veracity (the four V's) leading to big value [92], [93]. Accordingly, the large amount of heterogeneous data generated by healthcare and government

agencies needs to be exploited via proper big data analysis methods in order to improve personalized care, detection, treatment, prognosis and follow-up of various diseases. Service quality of healthcare centers, healthcare education, and even unemployment can be influenced by this [94].

Clinical benefits will likely require use of multi-modal data, as the influence of many factors (such as age [95]). Using images outside of their context seems like a task that is potentially too hard to be solved. Radiomics [96] allows extracting clinical markers from image data and the combination of this with genetic information can potentially lead to new and non-invasive methods, where benefits can become visible quickly.

## VII. OUTLOOK

Medical CBIR seems to be slower than retrieval in non-medical fields on several levels. Data sets have remained relatively small and annotated training examples are often available only in small amounts. This clearly creates many research opportunities. Generating shared, large data sets with annotations and running scientific challenges on them has much potential for real impact in identifying techniques that work in a stable way on varied data. It also seems important to use multi-modal data wherever possible as without the data that have an influence on the images it seems impossible to interpret the visual data. Thus, fusion techniques and ways to combine information from very different sources will be required in the future. This can help to analyze influences between data sources on the visual image data. For example, age influences visual information in medical images and specific diseases change patterns in other image parts.

The limited clinical use is another problem to be tackled. Maybe retrieval *per se* is not the most useful application and likely it needs to be integrated with applications of detection and classification to become useful in clinical practice and have a practical impact. The domain needs to be seen larger and include a detailed analysis of user requirements in specific clinical situations and for specific diseases such as cancer, instead of broad and general retrieval applications. With genomic, proteomic and metabolomic data and many other markers becoming available, the data integration challenge will increase in all fields of medicine and decision support tools that include visual data are required. Similar case retrieval is maybe more suited than similar image retrieval, meaning that again all data need to be integrated to understand what "similarity" in a specific situation actually means.

A current trend is the use of analysis and retrieval beyond radiology. Radiology data were the first to become digitally available in large quantities. Histopathology is starting to become digital and has many other challenges, with extremely large image sizes and a multi-scale analysis being usually necessary and with data driven approaches leading to extremely good results. Videos of endoscopy and ultrasound can likely have impact in clinical applications. Image and signal data have become available in many fields and need to be fully included into decision support.

One tendency that was found in this article is that less supervised approaches are being developed, which can leverage

large amounts of weakly annotated data or even data without annotations. These can likely be extracted automatically from hospital records and used for training algorithms but the limited veracity of the data need to be taken into account. Active learning can help to limit the annotation effort and concentrate on the most informative examples for manual annotation. Physicians can potentially be replaced in very focused tasks by laypersons using crowdsourcing but strict quality control is needed. We only found a few retrieval systems using features from deep learning and this will also likely explode in the coming years, as it did on other related areas of computer vision.

Rapid advances and developments in mobile health technologies require the adaptation of retrieval applications, which will allow access to clinical data and search functions in mobile scenarios [97]. This needs to be more integrated with global strategies [98] on the acquisition, storage and access to large scale distributed data sets. The socio-economic impact of these approaches should be evaluated better to really show the advantages and risks, and leverage on these.

The latter is linked to creating a real infrastructure for medical research data analysis and retrieval, potentially based on cloud techniques. Such an infrastructure could have all data remain at the producer's (i.e. hospital's) side and allow research to access the data for algorithm development and evaluation. This removes the barrier of data access, allows large-scale data analysis but requires data storage and computing infrastructures to be at the same place. EaaS is one concept in this area that could help the research community but solutions need to be developed for multi-institutional data analysis, where computation tasks carried out in different institutions can be combined in a central location.

All these tendencies and opportunities allow much research and the potential is high for real impact if integrated well with clinical applications that change quickly. Digital medicine is a reality and it is now important to develop quality tools for clinical impact of data integration that include visual data.

## REFERENCES

- [1] "Riding the wave: How Europe can gain from the rising tide of scientific data," Submission to the European Commission, available online at <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>, October 2010. [Online]. Available: <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>
- [2] C. Safran, M. Bloomrosen, W. E. Hammond, S. Labkoff, S. Markel-Fox, P. C. Tang, and D. E. Detmer, "Toward a national framework for the secondary use of health data: An American medical informatics association white paper," *MIM*, vol. 14, pp. 1–9, 2007.
- [3] A. Rosset, O. Ratib, A. Geissbuhler, and J.-P. Vallée, "Integration of a multimedia teaching and reference database in a PACS environment," *RG*, vol. 22, no. 6, pp. 1567–1577, December 2002.
- [4] L. H. Y. Tang, R. Hanka, and H. H. S. Ip, "A review of intelligent content-based indexing and browsing of medical images," *HII*, vol. 5, pp. 40–49, 1999.
- [5] 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*, vol. 73, no. 1, pp. 1–23, 2004.
- [6] C. Akgül, D. Rubin, S. Napel, C. Beaulieu, H. Greenspan, and B. Acar, "Content-based image retrieval in radiology: Current status and future directions," *Journal of Digital Imaging*, vol. 24, no. 2, pp. 208–222, 2011.
- [7] A. Kumar, J. Kim, W. Cai, M. Fulham, and D. Feng, "Content-based medical image retrieval: A survey of applications to multidimensional and multimodality data," *Journal of Digital Imaging*, vol. 26, no. 6, pp. 1025–1039, 2013.
- [8] A. M. Aisen, L. S. Broderick, H. Winer-Muram, C. E. Brodley, A. C. Kak, C. Pavlopoulou, J. Dy, C.-R. Shyu, and A. Marchiori, "Automated storage and retrieval of thin-section CT images to assist diagnosis: System description and preliminary assessment," *Radiology*, vol. 228, no. 1, pp. 265–270, July 2003.
- [9] D. Markonis, M. Holzer, F. Baroz, R. R. Ruiz de Castaneda, C. Boyer, G. Langa, H. Müller, and A. Depeursinge, "User-oriented evaluation of a medical image retrieval system for radiologists," *International journal of medical informatics*, 2015.
- [10] 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*, vol. 22, no. 12, pp. 1349–1380, December 2000.
- [11] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Computing Surveys*, vol. 40, no. 2, pp. 1–60, April 2008.
- [12] M. W. Vannier and R. M. Summers, "Sharing images," *Radiology*, vol. 228, pp. 23–25, 2003.
- [13] K. Doi, "Computer-aided diagnosis in medical imaging: Historical review, current status and future potential," *CMIG*, vol. 31, no. 4–5, pp. 198–211, mar 2007.
- [14] B. Elger, J. Iavindrasana, L. Lo Iacono, H. Müller, N. Roduit, P. Summers, and J. Wright, "Strategies for health data exchange for secondary, cross-institutional clinical research," *Computer Methods and Programs in Biomedicine*, vol. 99, no. 3, pp. 230–251, September 2010.
- [15] J. Hughes, D. Hunter, M. Sheehan, S. Wilkinson, and A. Wrigley, *European textbook on ethics in research*. Publications Office of the European Union, 2010.
- [16] D. Unay and A. Ekin, "Medical image search and retrieval for improved tele-healthcare," in *Telehealthcare Computing and Engineering*, F. Hu, Ed. CRC Press, 2013, ch. 22.
- [17] C. W. Cleverdon, "Report on the testing and analysis of an investigation into the comparative efficiency of indexing systems," Aslib Cranfield Research Project, Cranfield, USA, Tech. Rep., September 1962.
- [18] H. Müller, P. Clough, T. Deselaers, and B. Caputo, Eds., *ImageCLEF – Experimental Evaluation in Visual Information Retrieval*, ser. The Springer International Series On Information Retrieval. Springer, Berlin Heidelberg, 2010, vol. 32.
- [19] T. Heimann, B. Van Ginneken, M. Styner, Y. Arzhaeva, V. Aurich, C. Bauer, A. Beck, C. Becker, R. Beichel, G. Bekes *et al.*, "Comparison and evaluation of methods for liver segmentation from CT datasets," *Medical Imaging, IEEE Transactions on*, vol. 28, no. 8, pp. 1251–1265, 2009.
- [20] O. A. Jiménez del Toro, A. A. Taha, K. Gruenberg, M. Krenn, M. Winterstein, H. Müller *et al.*, "Cloud-based evaluation of organ segmentation and landmark detection algorithms: Visceral anatomy benchmarks," *IEEE Transactions on Medical Imaging*, 2016.
- [21] O. A. Jiménez del Toro, A. Foncubierta-Rodríguez, H. Müller, G. Langa, and A. Hanbury, "Overview of the visceral retrieval benchmark 2015," in *Multimodal Retrieval in the Medical Domain*, ser. Lecture Notes in Computer Science, vol. 9059. Springer, April 2015.
- [22] M. G. Ayadi, R. Bouslimi, and J. Akaichi, "A medical image retrieval scheme with relevance feedback through a medical social network," *Social Network Analysis and Mining*, vol. 6, no. 1, p. 53, 2016.
- [23] X. Zhang, H. Dou, T. Ju, J. Xu, and S. Zhang, "Fusing heterogeneous features from stacked sparse autoencoder for histopathological image analysis," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 5, pp. 1377–1383, 2016.
- [24] M. V. N. Bedo, D. Pereira dos Santos, M. Ponciano-Silva, P. M. de Azevedo-Marques, A. d. L. Ferreira de Carvalho, and C. Traina, "Endowing a content-based medical image retrieval system with perceptual similarity using ensemble strategy," *Journal of Digital Imaging*, vol. 29, no. 1, pp. 22–37, 2016.
- [25] A. Mustapha, A. Hussain, S. A. Samad, M. A. Zulkifley, W. M. Diyana Wan Zaki, and H. A. Hamid, "Design and development of a content-based medical image retrieval system for spine vertebrae irregularity," *BioMedical Engineering OnLine*, vol. 14, no. 1, p. 6, 2015.
- [26] M. Srinivas, R. R. Naidu, C. S. Sastry, and C. K. Mohan, "Content based medical image retrieval using dictionary learning," *Neurocomputing*, vol. 168, pp. 880–895, 11 2015.
- [27] M. Jiang, S. Zhang, J. Huang, L. Yang, and D. N. Metaxas, "Scalable histopathological image analysis via supervised hashing with multiple features," *Medical Image Analysis*, vol. 34, pp. 3–12, 12 2016.

- [28] X. Zhang, W. Liu, M. Dundar, S. Badve, and S. Zhang, "Towards large-scale histopathological image analysis: Hashing-based image retrieval," *IEEE Transactions on Medical Imaging*, vol. 34, no. 2, pp. 496–506, 2015.
- [29] H. C. Akakin and M. N. Gurcan, "Content-based microscopic image retrieval system for multi-image queries," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 4, pp. 758–769, 2012.
- [30] X. Zhou, R. Stern, and H. Müller, "Case-based fracture image retrieval," *International Journal of Computer Assisted Radiology and Surgery*, vol. 7, no. 3, pp. 401–411, 2012.
- [31] M. Jiang, S. Zhang, H. Li, and D. N. Metaxas, "Computer-aided diagnosis of mammographic masses using scalable image retrieval," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 2, pp. 783–792, 2015.
- [32] J. Liu, S. Zhang, W. Liu, C. Deng, Y. Zheng, and D. N. Metaxas, "Scalable mammogram retrieval using composite anchor graph hashing with iterative quantization," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. PP, no. 99, pp. 1–1, 2016.
- [33] Z. Çamlıca, H. R. Tizhoosh, and F. Khalvati, "Autoencoding the retrieval relevance of medical images," *2015 International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pp. 550–555, 2015.
- [34] X. Liu, H. R. Tizhoosh, and J. Kofman, "Generating binary tags for fast medical image retrieval based on convolutional nets and radon transform," *2016 International Joint Conference on Neural Networks (IJCNN)*, pp. 2872–2878, 2016.
- [35] M. K. Kundu, M. Chowdhury, and S. Das, "Interactive radiographic image retrieval system," *Computer Methods and Programs in Biomedicine*, vol. 139, pp. 209–220, 2 2017.
- [36] M. Babaie, H. R. Tizhoosh, S. Zhu, and M. E. Shiri, "Retrieving similar x-ray images from big image data using radon barcodes with single projections," *CoRR*, vol. abs/1701.00449, 2017.
- [37] R. Lan and Y. Zhou, "Medical image retrieval via histogram of compressed scattering coefficients," *IEEE Journal of Biomedical and Health Informatics*, vol. PP, no. 99, pp. 1–1, 2016.
- [38] Y. Anavi, I. Kogan, E. Gelbart, O. Geva, and H. Greenspan, "A comparative study for chest radiograph image retrieval using binary texture and deep learning classification," *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2940–2943, 2015.
- [39] A. Kumar, S. Dyer, J. Kim, C. Li, P. H. W. Leong, M. Fulham, and D. Feng, "Adapting content-based image retrieval techniques for the semantic annotation of medical images," *Computerized Medical Imaging and Graphics*, vol. 49, pp. 37–45, 4 2016.
- [40] G. Wei, H. Ma, W. Qian, and M. Qiu, "Similarity measurement of lung masses for medical image retrieval using kernel based semisupervised distance metric," *Medical Physics*, vol. 43, no. 12, pp. 6259–6269, 2016.
- [41] A. Shah, S. Conjeti, N. Navab, and A. Katouzian, "Deeply learnt hashing forests for content based image retrieval in prostate mr images," pp. 978 414–978 414–6, 2016.
- [42] S. Liu, S. Liu, W. Cai, H. Che, S. Pujol, R. Kikinis, M. Fulham, and D. Feng, "High-level feature based pet image retrieval with deep learning architecture," *Journal of Nuclear Medicine*, vol. 55, no. supplement 1, pp. 2028–2028, 05 2014.
- [43] S. Liu, S. Pujol, R. Kikinis, D. Feng, and W. Cai, "Propagation graph fusion for multi-modal medical content-based retrieval," *2014 13th International Conference on Control Automation Robotics & Vision (ICARCV)*, pp. 849–854, 2014.
- [44] Y. Zhuo, D. Ni, S. Chen, B. Lei, and T. Wang, "Adaptive ensemble manifold learning for neuroimaging retrieval," *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pp. 859–862, 2016.
- [45] J. R. F. Junior and M. C. Oliveira, "Evaluating margin sharpness analysis on similar pulmonary nodule retrieval," *2015 IEEE 28th International Symposium on Computer-Based Medical Systems*, pp. 60–65, 2015.
- [46] Y. Song, W. Cai, S. Eberl, M. J. Fulham, and D. Feng, "Thoracic image case retrieval with spatial and contextual information," *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 1885–1888, 2011.
- [47] W. Cai, Y. Song, and D. D. Feng, "Regression and classification based distance metric learning for medical image retrieval," *IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 1775–1778, 2012.
- [48] C. Beecks, K. Schoeffmann, M. Lux, M. S. Uysal, and T. Seidl, "Endoscopic video retrieval: A signature-based approach for linking endoscopic images with video segments," *2015 IEEE International Symposium on Multimedia (ISM)*, pp. 33–38, 2015.
- [49] J. R. Carlos, M. Lux, X. Giro-i Nieto, P. Munoz, and N. Anagnostopoulos, "Visual information retrieval in endoscopic video archives," *2015 13th International Workshop on Content-Based Multimedia Indexing (CBMI)*, pp. 1–6, 2015.
- [50] Z. Li, R. Fang, F. Shen, A. Katouzian, and S. Zhang, "Indexing and mining large-scale neuron databases using maximum inner product search," *Pattern Recognition*, vol. 63, pp. 680–688, 2017.
- [51] Y. Anavi, I. Kogan, E. Gelbart, O. Geva, and H. Greenspan, "Visualizing and enhancing a deep learning framework using patients age and gender for chest x-ray image retrieval," vol. 9785, 2016, pp. 978 510–978 510–6.
- [52] G. Quéllec, M. Lamard, G. Cazuguel, C. Roux, and B. Cochener, "Case retrieval in medical databases by fusing heterogeneous information," *IEEE Transactions on Medical Imaging*, vol. 30, no. 1, pp. 108–118, 2011.
- [53] A. Mourão, F. Martins, and J. Magalhães, "Multimodal medical information retrieval with unsupervised rank fusion," *Computerized Medical Imaging and Graphics*, vol. 39, pp. 35–45, 1 2015.
- [54] I. Kitanovski, G. Strezoski, I. Dimitrovski, G. Madjarov, and S. Loskovska, "Multimodal medical image retrieval system," *Multimedia Tools and Applications*, pp. 1–24, 2016.
- [55] J. Ramos, T. T. J. P. Kockelkorn, I. Ramos, R. Ramos, J. Grutters, M. A. Viergever, B. van Ginneken, and A. Campilho, "Content-based image retrieval by metric learning from radiology reports: Application to interstitial lung diseases," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 1, pp. 281–292, 2016.
- [56] F. Zhang, Y. Song, W. Cai, A. G. Hauptmann, S. Liu, S. Pujol, R. Kikinis, M. J. Fulham, D. D. Feng, and M. Chen, "Dictionary pruning with visual word significance for medical image retrieval," *Neurocomputing*, vol. 177, pp. 75–88, 2 2016.
- [57] M. Soydemir and D. Unay, "Context-aware medical image retrieval for improved dementia diagnosis," in *Intelligent Multimedia Technologies for Networking Applications: Techniques and Tools*, D. Kanellopoulos, Ed. Hershey, PA, USA: IGI Global, 2013, ch. 18, pp. 434–448.
- [58] A. García Seco de Herrera, J. Kalpathy-Cramer, D. Demner Fushman, S. Antani, and H. Müller, "Overview of the ImageCLEF 2013 medical tasks," in *Working Notes of CLEF 2013 (Cross Language Evaluation Forum)*, September 2013.
- [59] T. Tsirikika, A. García Seco de Herrera, and H. Müller, "Assessing the scholarly impact of ImageCLEF," in *CLEF 2011*, ser. Springer Lecture Notes in Computer Science (LNCS), sep 2011, pp. 95–106.
- [60] P. Clough, M. Sanderson, and H. Müller, "The CLEF Cross Language Image Retrieval Track (ImageCLEF) 2004," in *The Challenge of Image and Video Retrieval (CIVR 2004)*, ser. Lecture Notes in Computer Science (LNCS), vol. 3115. Springer, jul 2004, pp. 243–251.
- [61] A. F. Smeaton, W. Kraaij, and P. Over, "TRECVID 2003: An overview," in *Proceedings of the TRECVID 2003 conference*, December 2003.
- [62] M. Larson, G. J. Jones, B. Ionescu, M. Soleymani, and G. Gravier, "Recording and analyzing benchmarking results: The aims of the mediaeval working notes proceedings," in *MediaEval*, 2015.
- [63] B. H. Menze *et al.*, "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," *Medical Imaging, IEEE Transactions on*, vol. 34, no. 10, pp. 1993–2024, 2015.
- [64] A. Hanbury, H. Müller, G. Langs, M. A. Weber, B. H. Menze, and T. S. Fernandez, "Bringing the algorithms to the data: cloud-based benchmarking for medical image analysis," in *CLEF conference*, ser. Springer Lecture Notes in Computer Science, 2012.
- [65] O. Jimenez-del-Toro, A. Hanbury, G. Langs, A. Foncubierta-Rodríguez, and H. Müller, "Overview of the VISCERAL Retrieval Benchmark 2015," in *Multimodal Retrieval in the Medical Domain (MRMD) 2015*, ser. Lecture Notes in Computer Science, vol. 9059. Springer, 2015.
- [66] M. Krenn, M. Dorfer, O. A. Jiménez-del Toro, H. Müller, B. Menze, M.-A. Weber, A. Hanbury, and G. Langs, "Creating a Large-Scale Silver Corpus from Multiple Algorithmic Segmentations," in *MICCAI Workshop on Medical Computer Vision: Algorithms for Big Data, MCV 2015*, vol. 8848. Springer, 2014, pp. 163–170.
- [67] A. Hanbury, H. Müller, K. Balog, T. Brodt, G. V. Cormack, I. Eggel, T. Gollub, F. Hopfgartner, J. Kalpathy-Cramer, N. Kando, A. Krithara, J. Lin, S. Mercer, and M. Potthast, "Evaluation-as-a-service: Overview and outlook," *ArXiv*, vol. 1512.07454, 2015.
- [68] A. Foncubierta-Rodríguez and H. Müller, "Ground Truth Generation in Medical Imaging: A Crowdsourcing Based Iterative Approach," in *Workshop on Crowdsourcing for Multimedia*, oct 2012.
- [69] B. L. Ranard, Y. P. Ha, Z. F. Meisel, D. A. Asch, S. S. Hill, L. B. Becker, A. K. Seymour, and R. M. Merchant, "Crowdsourcing-harnessing the masses to advance health and medicine: a systematic review," *Journal of General Internal Medicine*, vol. 29, no. 1, pp. 187–203, 2014.
- [70] L. Maier-Hein, S. Mersmann, D. Kondermann, C. Stock, H. G. Kenngott, A. Sanchez, M. Wagner, A. Preukschas, A.-L. Wekerle, S. Helfert *et al.*, "Crowdsourcing for reference correspondence generation in endoscopic

- images,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2014, pp. 349–356.
- [71] D. Markonis, R. Donner, M. Holzer, T. Schlegl, R. Schaer, G. Langs, and H. Müller, “Khresmoi for radiologists - visual search in radiology archives and the open-access medical literature,” *Health Management*, vol. 13, no. 3, pp. 23–24, December 2013.
- [72] B. T. Wyman, D. J. Harvey, K. Crawford, M. A. Bernstein, O. Carmichael, P. E. Cole, P. K. Crane, C. DeCarli, N. C. Fox, J. L. Gunter *et al.*, “Standardization of analysis sets for reporting results from admi mri data,” *Alzheimer’s & Dementia*, vol. 9, no. 3, pp. 332–337, 2013.
- [73] J. Arevalo, A. Cruz-Roa *et al.*, “Histopathology image representation for automatic analysis: a state-of-the-art review,” *Revista Med*, vol. 22, no. 2, pp. 79–91, 2014.
- [74] C. B. Akgul, D. L. Rubin, S. Napel, C. F. Beaulieu, H. Greenspan, and B. Acar, “Content-based image retrieval in radiology: current status and future directions,” *J Digital Imaging*, vol. 24, 2011.
- [75] D. Unay, A. Ekin, and R. S. Jasinschi, “Local structure-based region-of-interest retrieval in brain mr images,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 4, pp. 897–903, 2010.
- [76] A. Depeursinge and H. Müller, “Fusion techniques for combining textual and visual information retrieval,” in *ImageCLEF*, ser. The Springer International Series On Information Retrieval, H. Müller, P. Clough, T. Deselaers, and B. Caputo, Eds. Springer Berlin Heidelberg, 2010, vol. 32, pp. 95–114.
- [77] Y. Bengio, “Deep learning and cultural evolution,” in *Genetic and Evolutionary Computation Conference, GECCO*, Vancouver, BC, Canada, 2014, pp. 1–2.
- [78] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A large-scale hierarchical image database,” in *IEEE Conference on Computer Vision and Pattern Recognition*, ser. CVPR 2009, 2009, pp. 248–255.
- [79] G. J. S. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghahfoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, “A survey on deep learning in medical image analysis,” *CoRR*, vol. abs/1702.05747, 2017.
- [80] H. Greenspan, B. van Ginneken, and R. M. Summers, “Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1153–1159, 2016.
- [81] H. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. J. Mollura, and R. M. Summers, “Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning,” *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.
- [82] J. Cheng, D. Ni, Y. Chou, J. Qin, C. Tiu, Y. Chang, C. Huang, D. Shen, and C. Chen, “Computer-aided diagnosis with deep learning architecture: Applications to breast lesions in us images and pulmonary nodules in ct scans,” *Scientific Reports*, vol. 6, 4 2016.
- [83] A. Janowczyk and A. Madabhushi, “Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases,” *Journal of Pathology Informatics*, vol. 7, 2016.
- [84] A. A. Cruz-Roa, J. E. A. Ovalle, A. Madabhushi, and F. A. G. Osorio, “A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2013, pp. 403–410.
- [85] P. Groves, B. Kayyali, D. Knott, and S. V. Kuiken, “The ‘big data’ revolution in healthcare: Accelerating value and innovation,” McKinsey & Company, Center for US Health System Reform, Business Technology Office, Tech. Rep., January 2013.
- [86] A. M. Rutman and M. D. Kuo, “Radiogenomics: Creating a link between molecular diagnostics and diagnostic imaging,” *European journal of radiology*, vol. 70, no. 2, pp. 232–241, May 2009.
- [87] J. P. A. Ioanidis, “Why most published research findings are false,” *PLoS One*, vol. 2, no. 8, 2005.
- [88] I. U. Khan, S. ur Rehman *et al.*, “A review on big data security and privacy in healthcare applications,” in *Big Data Management*. Springer, 2017, pp. 71–89.
- [89] W. Raghupathi and V. Raghupathi, “Big data analytics in healthcare: promise and potential,” *Health Information Science and Systems*, vol. 2, p. 3, 2014.
- [90] H. K. Patil and R. Seshadri, “Big data security and privacy issues in healthcare,” *2014 IEEE International Congress on Big Data*, pp. 762–765, 2014.
- [91] O. Tene and J. Polonetsky, “Big data for all: Privacy and user control in the age of analytics,” *Northwestern Journal of Technology and Intellectual Property*, vol. 11, no. 5, pp. 239–273, 2013.
- [92] A. Belle, R. Thiagarajan, S. M. R. Soroushmehr, F. Navidi, D. A. Beard, and K. Najarian, “Big data analytics in healthcare,” *BioMed Research International*, vol. 2015, p. 16, 2015.
- [93] T. Huang, L. Lan, X. Fang, P. An, J. Min, and F. Wang, “Promises and challenges of big data computing in health sciences,” *Big Data Research*, vol. 2, no. 1, pp. 2 – 11, 2015, special Issue on Computation, Business, and Health Science.
- [94] J. Archena and E. M. Anita, “A survey of big data analytics in healthcare and government,” *Procedia Computer Science*, vol. 50, pp. 408 – 413, 2015.
- [95] Y. Dicente Cid, A. Mamonov, A. Beers, A. Thomas, V. Kovalev, J. Kalpathy-Cramer, and H. Müller, “Making sense of large datasets without annotations: Analyzing age-related correlations from lung CT scans,” in *SPIE Medical Imaging*. International Society for Optics and Photonics, 2017.
- [96] S. Napel and M. Giger, “Special section guest editorial: radiomics and imaging genomics: Quantitative imaging for precision medicine,” *Journal of Medical Imaging*, vol. 2, no. 4, p. 041001, 2015.
- [97] A. Depeursinge, S. Duc, I. Eggel, and H. Müller, “Mobile medical visual information retrieval,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 1, pp. 53–61, January 2012.
- [98] T. Group, “The socio-economic impact of mobile health,” April 2012. [Online]. Available: <https://www.telenor.com/wp-content/uploads/2012/05/BCG-Telenor-Mobile-Health-Report-May-20121.pdf>



**Henning Müller** studied medical informatics at the University of Heidelberg, Germany, then worked at Daimler-Benz research in Portland, OR, USA. From 1998-2002 he worked on his PhD degree at the University of Geneva, Switzerland with a research stay at Monash University, Melbourne, Australia in 2001. Since 2002 Henning has been working in medical informatics at the University Hospitals of Geneva where he habilitated in 2008 and was named titular professor in 2014. Since 2007 he has been a professor in business informatics at the HES-SO Valais in Sierre, Switzerland and since 2011 responsible for the eHealth unit. Henning was coordinator of the Khresmoi project, scientific coordinator of VISCERAL, initiator of the ImageCLEF benchmark. He has authored over 400 scientific papers, is in the editorial board of several journals and reviews for many journals and funding agencies. In 2015–2016 Henning was visiting professor at the Martinos Center in Boston, MA, USA part of Harvard Medical School and the Massachusetts General Hospital (MGH) working on collaborative projects in medical image analysis and system evaluation in the context of the Quantitative Imaging Network of the National Cancer Institute.



**Devrim Ünay** received the Ph.D. degree in applied sciences from Faculty of Engineering Mons, Belgium in 2006. He then was with the Video Processing and Analysis Group of Philips Research, Netherlands as a Marie-Curie fellow (2007-2009) and with the VPA Lab of Sabanci University, Turkey (2009-2010). From 2010 to 2015, he served as a faculty member at Bahcesehir University, Turkey.

Since 2015, he has been a faculty member at Izmir University Economics, Turkey as associate professor. His current research interests include image processing, computer vision, pattern recognition and information retrieval with a focus on applications in biological and medical image analysis. He has been the principal investigator of 4 national and 1 international projects. He has authored over 90 scientific papers, holds two US patents, has served as program/organizing committee member of several international conferences, and acts as a reviewer for many journals and funding agencies.