

Studying Public Medical Images from the Open Access Literature and Social Networks for Model Training and Knowledge Extraction

Henning Müller^{1,2}, Vincent Andrearczyk¹, Oscar Jimenez del Toro¹, Anjani Dhrangadhariya¹, Roger Schaefer¹, and Manfredo Atzori¹

¹ University of Applied Sciences Western Switzerland (HES-SO), Sierre, Switzerland

² Medical faculty, University of Geneva, Geneva, Switzerland

henning.mueller@hevs.ch

Abstract. Medical imaging research has long suffered problems getting access to large collections of images due to privacy constraints and to high costs that annotating images by physicians causes. With public scientific challenges and funding agencies fostering data sharing, repositories, particularly on cancer research in the US, are becoming available. Still, data and annotations are most often available on narrow domains and specific tasks. The medical literature (particularly articles contained in MedLine) has been used for research for many years as it contains a large amount of medical knowledge. Most analyses have focused on text, for example creating semi-automated systematic reviews, aggregating content on specific genes and their functions, or allowing for information retrieval to access specific content. The amount of research on images from the medical literature has been more limited, as MedLine abstracts are available publicly but no images are included. With PubMed Central, all the biomedical open access literature has become accessible for analysis, with images and text in structured format. This makes the use of such data easier than extracting it from PDF. This article reviews existing work on analyzing images from the biomedical literature and develops ideas on how such images can become useful and usable for a variety of tasks, including finding visual evidence for rare or unusual cases. These resources offer possibilities to train machine learning tools, increasing the diversity of available data and thus possibly the robustness of the classifiers. Examples with histopathology data available on Twitter already show promising possibilities. This article add links to other sources that are accessible, for example via the ImageCLEF challenges.

Keywords: medical imaging, biomedical literature, machine learning, training

1 Introduction

Machine learning, in particular deep learning, relies on large annotated datasets to reach sometimes impressive results, as for example in the ImageNet competition [9]. Whereas web images are easily available and annotations can be

obtained via crowdsourcing without a need for expensive specialists, the medical domain suffers more difficulties in creating large-scale image resources that are publicly available. All medical data acquisitions require ethics approval and privacy constraints can sometimes limit the distribution and sharing of such data even when it is considered of high importance [39]. As a consequence, small and non publicly available datasets limit the use of state-of-the-art machine learning algorithms that require large amounts of data, as well as the reproducibility of many reported results in the literature. Many funding organizations in the US and Europe have started pushing for data sharing and for making resources available over 15 years ago. The outcomes are now visible with large scale repositories such as the TCIA³ (The Cancer Imaging Archive) and TCGA⁴ (The Cancer Genome Atlas) being increasingly used by research projects. These repositories take some of the legal responsibilities from the medical institutions that acquired that data, easing data sharing for them. Scientific challenges have also contributed to making medical data available for concrete tasks with strong baselines. ImageCLEF [20, 11] has had medical challenges each year since 2004 and, together with other challenges such as VISCERAL [18], BraTS [28] or Camelyon [4], has strongly gained in popularity. Even though scientific challenges have strongly advanced many domains, there is a risk of bias (depending on the exact setup and performance measures used) and results need to be evaluated with care [25]. Other resources for access to images and videos include [32, 34].

The biomedical literature stores a very large amount of available medical knowledge. Medical knowledge is estimated to have a half-life of about 5-10 years [13], meaning that knowledge needs to be checked regularly and it may be important to use recent data to assure veracity when training systems. Similarly for images, the acquisition devices change continuously. Ways to harmonize images across the imaging devices also need to be developed [1].

While a large amount of work has been dedicated to the automatic analysis of text in biomedical literature [40, 37], work on extracting and practically using images present in the literature is still limited. A major difficulty that explains this limited amount of work is the unstructured nature of the data, with images not being easily extracted from the literature and associated with meaningful information or labels. The biomedical literature also contains an extreme variability of information, making it necessary to filter out the unwanted data [20].

The PubMed Central (PMC) repository⁵ has allowed this type of analysis in recent years, since an exponentially increasing number of images has become available with the free text and limited structured data (such as attached MeSH terms, keywords, place where a figure is referenced, etc.) (see Figure 1). With more than 2 million articles in total, an average of 3.5 figures per article including 1.5 compound figures of 4 subfigures each, it sums up to approximately seven million figures, including 3 million compound figures with 12 million subfigures for a total of >16 million figures available in 2018 if separated correctly. With

³ <https://www.cancerimagingarchive.net/>

⁴ <https://www.genome.gov/Funded-Programs-Projects/Cancer-Genome-Atlas>

⁵ <http://www.ncbi.nlm.nih.gov/pmc/>

an expected increase of nearly 3 million figures in 2019 and more in the following years, it promises in the near future very large amounts of training data in various applications, modalities and particularly rare cases that are strongly oversampled in the literature compared to clinical archives [12]. Difficulties related to this type of data, that will be developed in more details in Section 3, include the heterogeneity and non-guaranteed quality of the images, the presence of compound images and the automation of ground truth labels extraction from the text. This article has as objective the systematic analysis of work towards

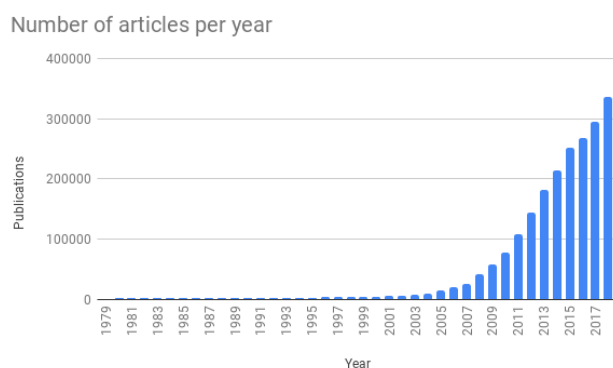


Fig. 1. The growth in the number of published open access articles in the biomedical literature has been exponential in the past 30 years.

using publicly available sources of medical images, such as PMC. It reviews much of the existing work in the field of making such resources usable for machine learning. The main contribution of this text is the systematic collection of available articles in the field including resources of annotated data sets that can be used for model training. Thus, this article makes the available resources (articles, data sets, source code) accessible for researchers, resources that are currently scattered and not easy to find. It solves the problem of where to start when filtering the medical literature for relevant images for a specific problem. The example problem chosen here is that of histopathology image analysis.

2 Methods

For gathering the articles cited in this text, two approaches were chosen as a more qualitative methodology. As a first approach, the work of the ImageCLEF benchmark [20] that has extensively worked on images from the biomedical open-access literature was analyzed over the past 12 years, a period in which most image resources have come from the literature or other public sources. Several steps were analysed in this context to extract information from images in the

literature, to filter our relevant clinical images and to retrieve visual content or answer questions based on this visual content. In a second step, the literature search tool Google scholar was used to add articles working on related fields and for specific tasks. Search terms include “medical image classification” and “medical literature” and “publicly available resources” and “machine learning”. Articles found were chosen based on being complementary to the already cited articles, journal papers were favored over conference papers and dynamically growing data sets (such as PMC) were favored over fixed data sets. The found resources are sorted into several categories based on the steps required for the data extraction and data enrichment.

3 Analyzing Images from the Biomedical Literature

3.1 Retrieval of Medical Images

One of the first ways to exploit images from the biomedical literature was image retrieval systems, that indexed images and allowed to search by keywords and visual examples or for visually similar regions of interest. This made the visual content accessible and allowed to reuse single images with specific patterns or groups of images more easily. The OpenI research system [8], maintained by the National Library of Medicine (NLM) and shown in Figure 2 is an example of such retrieval systems. A similar example with more detailed visual search

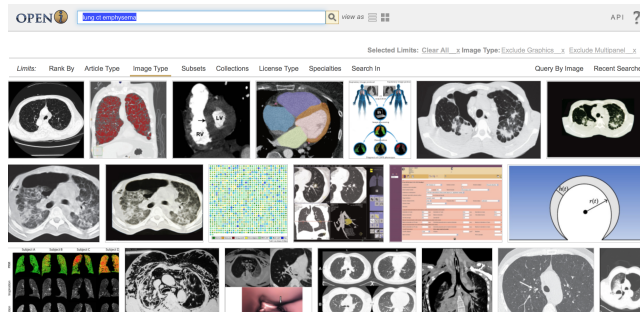


Fig. 2. Screenshot of the OpenI image retrieval system [8].

capabilities was developed in the Khresmoi project [26] based on [30], while another one focusing on radiology without visual search is GoldMiner [19]. A large number of further medical retrieval systems exist as reviewed in [31].

3.2 Extracting Content from Medical Images

When the objectives go beyond finding specific images and require filtering the entire PMC, then it is necessary to extract metadata from the images, which is described in the following sections.

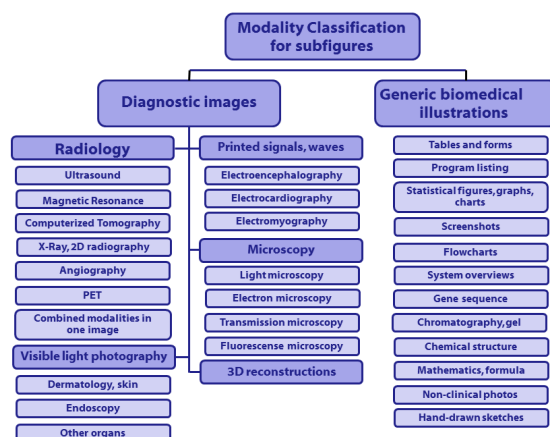


Fig. 3. A hierarchy of images types in the biomedical literature, taken from [29].

Image Types The first basic information to use for filtering an image collection is the image type, for example, the imaging modality that produced the image but also whether a flow chart or other general graph is shown. Image search engines such as GoldMiner [19], for instance, allow to filter a search by image modality and it was shown that filtering by modality can improve retrieval quality [21]. Specifically for the biomedical literature in PMC, a hierarchy of image types was developed in [29]. This hierarchical structure is also shown in Figure 3. Several challenges on identifying the image types were run in ImageCLEF [15]. The currently best results for the really challenging and very unbalanced data set of over 30 classes reached over 90% [2].

In order to underline the diversity in the image types, we show in Figure 4 a set of images that were initially classified as histopathology images but in manual control found to be different types. These examples illustrate the visual similarity of some image types and, in one case, a hand-drawn histopathology image which clearly looks like histopathology but is a different type in the hierarchy. Besides the figure type that was also described in the IRMA (Image Retrieval in Medical Applications) hierarchy [22] (that focuses only on medical modalities) there is also an interest to identify the anatomical region of the images, its orientation and the biological system that is imaged (the heart vs. the lung for a chest X-ray, for example). These meta-data allow for good filtering and search. Pathology is the most frequently requested functionality for search [27] but it also very to obtain from the image data alone, as often more information on a case is required.

Dealing with Compound Figures. The percentage of compound figures (consisting of several subfigures) is estimated to be at least 40% in PMC [15]. This high ratio can be explained by the fact that many journals allow for a very limited

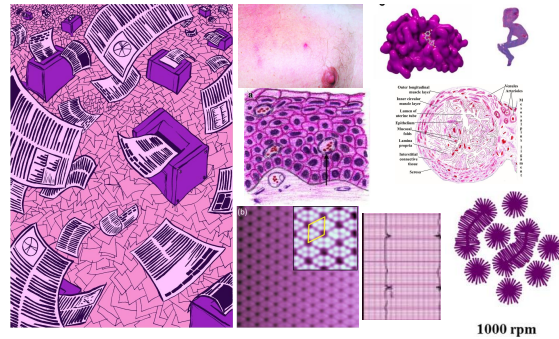


Fig. 4. Examples of images incorrectly classified as histopathology images and manually identified as incorrect categories. Histopathology images can be seen in Figure 5.

number of figures, forcing to compress the visual content. These images are only available in a single image block limiting their utility in computer analysis or for further analysis. Making the individual subfigures accessible requires detecting the compound figures among the millions of figures and splitting them into subfigures as explained in [3, 6]. Several challenges arise, as the subfigures are often related and sometimes it is even hard to identify whether a figure is a single figure or contains independent subparts. An example of journal compound figures with several subfigures that are automatically separated is shown in Figure 5, in this case focusing on histopathology images.

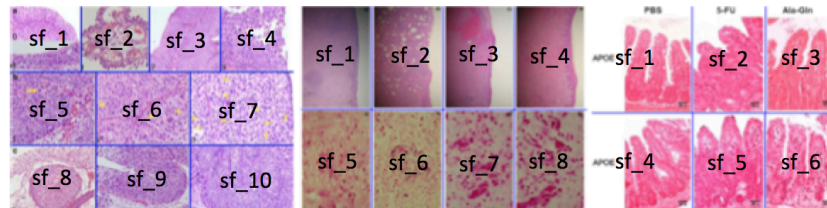


Fig. 5. Examples of compound figures and the automatically detected separation lines.

As with the image types described before, the variety of images and ways in which images are put into compound figures is enormous. Again, several challenges in ImageCLEF tackled this problem with often very good results for both detecting the subfigures and separating them [16]. After subfigures are cut they can again be classified into image types and then be used for specific tasks.

Focusing on Regions of Interest. In medical images, the actual region of interest is often small. In journal articles, the raw data are usually not made available and replaced by cropped images that are already in the optimal level/window setting for viewing. These images show a subset of the original image and for

tomographic images most frequently only the best slides, for example where a tumor is largest. This reduction of information, intrinsic to the purpose of reporting images in scientific articles, presents inconveniences. Information is lost with the compression and transfer to jpeg format (256 grey levels instead of 4000 in DICOM) and with the omission of slides in tomographic series. Yet, the content is also more focused and often shows only the environment around the region of interest. This can possibly allow weakly annotated approaches if a sufficiently large amount of data is available. Another particularity of images from the literature is the presence of arrows to highlight regions that are further discussed in the caption or the full text [5]. These arrows can relatively easily be detected and allow to further focus the search on small regions of interest [5].

3.3 Combining Text and Images for Data Analysis

Much of the description in the preceding sections has focused on image analysis approaches to enrich the information of images from the medical literature by filtering specific image types, cutting compound figures and detecting the region of interest. In the case of the biomedical literature, there is also text information available, notably the figure caption and the full text surrounding the reference to a figure. In PMC, most articles have manually attached MeSH (Medical Subject Headings) terms. Text has been used in most retrieval applications [10] but has also obtained very good results in modality classification [17, 2], as it is complementary to visual information. In the case of compound figure detection, a caption with several subparts can also be indicative for the presence of subfigures. The text can then be split to correspond to each of the subfigures [3].

For images detected as histopathology to be useful for training clinical decision support systems, further steps are required [12]. Much of the cancer research is on animals so, when learning human tissue classification, it is important to filter out animal tissue samples even from the same organ. The species is most often present in the manually attached MeSH terms (for humans even the group of children, or older adults is usually given) and these can be used as a good indicator for filtering out those of mice and rats. The specific organ is often mentioned in the MeSH terms or in the article title but is only rarely mentioned in the figure caption. This again allows filtering out tissue of organs that are not relevant for a given task.

Image Heterogeneity in the Case of Histopathology. Histopathology has several specificities that increase the possible image heterogeneity. Different types of stainings can highlight various molecular aspects in the images in differing colors. The most frequent one, the H&E (Hematoxylin and Eosin) staining, is low in cost and accounts for approximately 90% of the histopathology images in the literature based on a small sample that we evaluated. Within H&E staining, color heterogeneity can strongly affect the performance of the machine learning algorithms [7], so heterogeneity needs to be added in the training data. In pathology images, color variations can be due to tissue preparation and are related to the complex set of preparation phases related to staining procedures, section

thickness and scanner differences [23]. Color heterogeneity can be approached with many methods, such as color normalization [38, 24, 7] and color augmentation [36] directly in the learning phase. Particularly, color augmentation has recently lead to excellent results in challenges [7]. Interpretability methods can evaluate the impact of color shifts and texture transformations [14]. Besides this, detecting the scale of figures [33] is another challenge. Structures being analyzed at differing scales on a microscope, it is important to know the image scale when comparing images or when using them for machine learning. Sometimes, this information can be found in the image caption but it may not actually be reliable, as the editors can modify the image size and resolution.

3.4 An Example of Using Twitter Histopathology Images

Images and information posted by pathologists on social media (in this case Twitter) were used in [35] to create a dataset and train machine learning algorithms to identify stains and discriminate between different tissues in histopathology images. The results obtained by Schaumberg et al. [35] show that such data can actually be used. However, they also highlight the limitations of the approach. Despite having a large number of manually curated images, good performance was only obtained for tasks that can be defined as very simple for a human. For instance, the authors reported Area Under Receiver Operating Characteristic (AUROC) over 0.9 for tasks such as: differentiating between human H&E stained microscopy images from all other types of images or distinguishing H&E from immunohistochemistry (IHC) stained microscopy images, a very simple task. AUROC reduces to 0.803 when distinguishing among breast, dermatological, gastrointestinal, genitourinary and gynecological pathology tissue types. Finally, the performance for more relevant tasks noticeably drops. Distinguishing between low-grade and malignant tumors on the mentioned tissues obtains an average AUROC of 0.703, while the three-class classification of nontumoral diseases, low-grade tumors, and malignant tumors drops to an average AUROC of 0.683. Both results should be considered in the light of the considered number of classes (respectively 2 and 3), which leads to chance level accuracies of respectively 50% and 33.3% in case of a balanced distribution.

4 Conclusions

The objective of this article is to gather and review research that aims at leveraging the usefulness of medical images publicly available on the Internet via a variety of sources. The most frequently used resource is the biomedical literature, although some communities shared images on social networks such as Twitter and there is potentially useful content in public teaching files and possibly other resources. With an exponentially increasing amount of images, it is important to develop automatic pipelines that adapt to the changing images and regularly add such up-to-date content to learn from it and really optimize generalization performance of classifiers with limited manual work. Many of the examples

shown in this paper focused on basic classification and filtering tasks to augment the value of the extracted images. This allowed, in internal experiments, to find around 100,000 histopathology images among the five million images in PMC (a number that is increasing very quickly). This number gets further reduced when removing non-human tissue and also when focusing on specific organs such as prostate or breast tissue.

Still, such teaching files, social networks or literature can provide important complements to clinical archives if the information contained in the images can be accurately extracted. Moreover, these resources focus on rare cases or abnormal situations and can complement clinical archives where the most common conditions are usually oversampled and rare or unusual cases are not often seen.

This paper focused on a scenario using histopathology images and it shows the steps of how to make images from the biomedical literature usable for machine learning tasks. A quantitative evaluation is clearly required to show the usefulness of these images to machine learning. A very similar approach can be employed in radiology or dermatology, simply filtering for a different set of images and different subclasses. Each time, resources can be created that have few problems in terms of privacy constraints and ethics because the data have already been made available in the biomedical literature with a clear license. The work required can be important in a first step to create a fully automated pipeline but it is still much less than extracting images from clinical archives. The quickly increasing data source rewards with the possibility to have within a few years a much larger resource available and being able to reuse such an automatic pipeline to retrain the system.

Using images from the literature is complementary to using images from clinical archives for training, as clinical archives focus on local specificities whereas images in the literature allow for better generalization. Images from the literature contain mainly unusual or rare cases whereas images in clinical archives contain many normal or frequent cases. Thus, the two sources can be used together. Finally, the literature also allows groups without connections to a medical center to work on medical data. The next steps in this work are to show that training with such heterogeneous images are possible for a well-defined task such as grading of prostate cancer histopathology images. Other useful targets would be to create an automated pipeline for all the steps required to enrich images from the literature and then run this at several moments over time to estimate data increase. If several research groups can collaborate, a possible critical mass can be reached and combinations of automatic labels could reach a high quality. It would be even more useful if information on image type, compound figures etc. could be distributed directly with PMC, so all research using the data can profit from it. In the case of uncertain labels, a confidence score could be added.

5 Acknowledgment

This work was partially funded by the EU H2020 ExaMode H2020 project (grant agreement 825292).

References

1. Andrearczyk, V., Depeursinge, A., Müller, H.: Neural network training for cross-protocol radiomic feature standardization in computed tomography. *Journal of Medical Imaging* **6**(2), 024008 (2019)
2. Andrearczyk, V., Müller, H.: Deep multimodal classification of image types in biomedical journal figures. In: *International Conference of the Cross-Language Evaluation Forum for European Languages*. pp. 3–14. Springer (2018)
3. Apostolova, E., You, D., Xue, Z., Antani, S., Demner-Fushman, D., Thoma, G.R.: Image retrieval from scientific publications: text and image content processing to separate multi-panel figures. *Journal of the American Society for Information Science* **64**, 893–908 (2013)
4. Bejnordi, B.E., Veta, M., Van Diest, P.J., Van Ginneken, B., Karssemeijer, N., Litjens, G., Van Der Laak, J.A., Hermsen, M., Manson, Q.F., Balkenhol, M., et al.: Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *Journal of the American Medical Association* **318**(22), 2199–2210 (2017)
5. Cheng, B., Stanley, R.J., De, S., Antani, S., Thoma, G.R.: Automatic detection of arrow annotation overlays in biomedical images. *International Journal of Healthcare Information Systems and Informatics* **6**(4), 23–41 (2011)
6. Chhatkuli, A., Markonis, D., Foncubierta-Rodríguez, A., Meriaudeau, F., Müller, H.: Separating compound figures in journal articles to allow for subfigure classification. In: *SPIE Medical Imaging* (2013)
7. Cruz-Roa, A., Gilmore, H., Basavanthally, A., Feldman, M., Ganesan, S., Shih, N.N., Tomaszewski, J., González, F.A., Madabhushi, A.: Accurate and reproducible invasive breast cancer detection in whole-slide images: A deep learning approach for quantifying tumor extent. *Scientific Reports* **7**, 46450 (2017)
8. Demner-Fushman, D., Antani, S., Simpson, M.S., Thoma, G.R.: Design and development of a multimodal biomedical information retrieval system. *Journal of Computing Science and Engineering* **6**(2), 168–177 (2012)
9. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: A large-scale hierarchical image database. In: *IEEE Conference on Computer Vision and Pattern Recognition*. pp. 248–255. CVPR 2009 (2009)
10. Depeursinge, A., Müller, H.: Sensors, medical images and signal processing: Comprehensive multi-modal diagnosis aid frameworks. *IMIA Yearbook of Medical Informatics* **5**(1), 43–46 (2010)
11. Deselaers, T., Deserno, T.M., Müller, H.: Automatic medical image annotation in ImageCLEF 2007: Overview, results, and discussion. *Pattern Recognition Letters* **29**(15), 1988–1995 (2008)
12. Dhrangadhariya, A.K., Jimenez-del Toro, O., Andrearczyk, V., Atzori, M., Müller, H.: Exploiting the pubmed central repository to mine out a large multimodal dataset of rare cancer studies. In: *SPIE. International Society for Optics and Photonics* (2020)
13. Emanuel, E.: A half-life of 5 years. *Canadian Medical Association Journal* **112**(5), 572 (1975)
14. Graziani, M., Andrearczyk, V., Müller, H.: Regression concept vectors for bidirectional explanations in histopathology. In: *Understanding and Interpreting Machine Learning in Medical Image Computing Applications*, pp. 124–132. Springer (2018)
15. García Seco de Herrera, A., Kalpathy-Cramer, J., Demner Fushman, D., Antani, S., Müller, H.: Overview of the ImageCLEF 2013 medical tasks. In: *Working Notes of CLEF 2013 (Cross Language Evaluation Forum)* (September 2013)

16. García Seco de Herrera, A., Müller, H., Bromuri, S.: Overview of the ImageCLEF 2015 medical classification task. In: Working Notes of CLEF 2015 (Cross Language Evaluation Forum) (September 2015)
17. García Seco de Herrera, A., Schaer, R., Bromuri, S., Müller, H.: Overview of the ImageCLEF 2016 medical task. In: Working Notes of CLEF 2016 (Cross Language Evaluation Forum) (September 2016)
18. Jimenez-del-Toro, O., Müller, H., Krenn, M., Gruenberg, K., Taha, A.A., Winterstein, M., Eggel, I., Foncubierta-Rodríguez, A., Goksel, O., Jakab, A., Kontokotsios, G., Langs, G., Menze, B., Salas Fernandez, T., Schaer, R., Walleyo, A., Weber, M.A., Dicente Cid, Y., Gass, T., Heinrich, M., Jia, F., Kahl, F., Kechichian, R., Mai, D., Spanier, A.B., Vincent, G., Wang, C., Wyeth, D., Hanbury, A.: Cloud-based evaluation of anatomical structure segmentation and landmark detection algorithms: VISCERAL Anatomy Benchmarks. *IEEE Transactions on Medical Imaging* **35**(11), 2459–2475 (2016)
19. Kahn Jr., C.E., Thao, C.: Goldminer: A radiology image search engine. *American Journal of Roentgenology* **188**, 1475–1478 (2008)
20. Kalpathy-Cramer, J., García Seco de Herrera, A., Demner-Fushman, D., Antani, S., Bedrick, S., Müller, H.: Evaluating performance of biomedical image retrieval systems: Overview of the medical image retrieval task at ImageCLEF 2004–2014. *Computerized Medical Imaging and Graphics* **39**(0), 55 – 61 (2015)
21. Kalpathy-Cramer, J., Hersh, W.: Automatic image modality based classification and annotation to improve medical image retrieval. *Studies in Health Technology and Informatics* **129**, 1334–1338 (2007)
22. Lehmann, T.M., Schubert, H., Keyzers, D., Kohnen, M., Wein, B.B.: The IRMA code for unique classification of medical images. In: Huang, H.K., Ratib, O.M. (eds.) *Medical Imaging 2003: PACS and Integrated Medical Information Systems: Design and Evaluation*. SPIEProc, vol. 5033, pp. 440–451. San Diego, California, USA (May 2003)
23. Leo, P., Lee, G., Shih, N.N.C., Elliott, R., Feldman, M.D., Madabhushi, A.: Evaluating stability of histomorphometric features across scanner and staining variations: prostate cancer diagnosis from whole slide images. *Journal of Medical Imaging* **3**(4), 047502–047502 (2016)
24. Li, X., Plataniotis, K.N.: A complete color normalization approach to histopathology images using color cues computed from saturation-weighted statistics. *IEEE Transactions on Biomedical Engineering* **62**(7), 1862–1873 (2015)
25. Maier-Hein, L., Eisenmann, M., Reinke, A., Onogur, S., Stankovic, M., Scholz, P., Arbel, T., Bogunovic, H., Bradley, A.P., Carass, A.e.a.: Why rankings of biomedical image analysis competitions should be interpreted with care. *Nature communications* **9**(1), 5217 (2018)
26. Markonis, D., Donner, R., Holzer, M., Schlegl, T., Schaer, R., Langs, G., Müller, H.: Khresmoi for radiologists - visual search in radiology archives and the open-access medical literature. *Health Management* **13**(3), 23–24 (Dec 2013)
27. Markonis, D., Holzer, M., Dungs, S., Vargas, A., Langs, G., Kriewel, S., Müller, H.: A survey on visual information search behavior and requirements of radiologists. *Methods of Information in Medicine* **51**(6), 539–548 (2012)
28. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Criminisi, A., Das, T., Delingette, H., Demiralp, C., Durst, C.R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P.,

- Guo, X., Hamamci, A., Iftekharuddin, K.M., Jena, R., John, N.M., Konukoglu, E., Lashkari, D., Mariz, J.A., Meier, R., Pereira, S., Precup, D., Price, S.J., Raviv, T.R., Reza, S.M.S., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.C., Shotton, J., Silva, C.A., Sousa, N., Subbanna, N.K., Szekely, G., Taylor, T.J., Thomas, O.M., Tustison, N.J., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Van Leemput, K.: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *Medical Imaging, IEEE Transactions on* **34**(10), 1993–2024 (2015)
29. Müller, H., Kalpathy-Cramer, J., Demner-Fushman, D., Antani, S.: Creating a classification of image types in the medical literature for visual categorization. In: *SPIE Medical Imaging* (2012)
 30. Müller, H., Rosset, A., Vallée, J.P., Geissbuhler, A.: Integrating content-based visual access methods into a medical case database. In: *Proceedings of the Medical Informatics Europe Conference (MIE 2003)*. vol. 95, pp. 480–485. St. Malo, France (May 2003)
 31. Müller, H., Unay, D.: Retrieval from and understanding of large-scale multi-modal medical datasets: A review. *IEEE transactions on multimedia* **19**(9), 2093–2104 (2017)
 32. Münzer, B., Schoeffmann, K., Böszörményi, L.: Content-based processing and analysis of endoscopic images and videos: A survey. *Multimedia Tools and Applications* **77**(1), 1323–1362 (2018)
 33. Otálora, S., Atzori, M., Andrearczyk, V., Müller, H.: Image magnification regression using densenet for exploiting histopathology open access content. In: *Computational pathology and ophthalmic medical image analysis*, pp. 148–155. Springer (2018)
 34. Pogorelov, K., Randel, K.R., Griwodz, C., Eskeland, S.L., de Lange, T., Johansen, D., Spampinato, C., Dang-Nguyen, D.T., Lux, M., Schmidt, P.T.e.a.: Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection. In: *Proceedings of the 8th ACM on Multimedia Systems Conference*. pp. 164–169. ACM (2017)
 35. Schaumberg, A.J., Juarez, W., Choudhury, S.J., Pastrian, L.G., Pritt, B.S., Pozuelo, M.P., Sanchez, R.S., Ho, K., Zahra, N., Sener, B.D.: Large-scale annotation of histopathology images from social media. *bioRxiv* p. 396663 (2018)
 36. Tellez, D., Litjens, G., Bandi, P., Bulten, W., Bokhorst, J.M., Ciompi, F., van der Laak, J.: Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. *arXiv preprint arXiv:1902.06543* (2019)
 37. Tsatsaronis, G., Balikas, G., Malakasiotis, P., Partalas, I., Zschunke, M., Alvers, M.R., Weissenborn, D., Krithara, A., Petridis, S., Polychronopoulos, D., et al.: An overview of the bioasq large-scale biomedical semantic indexing and question answering competition. *BMC bioinformatics* **16**(1), 138 (2015)
 38. Valavanis, L., Stathopoulos, S.: IPL at ImageCLEF 2017 concept detection task. In: *CLEF2017 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org* <<http://ceur-ws.org>>, Dublin, Ireland (September 11-14 2017)
 39. Vannier, M.W., Summers, R.M.: Sharing images. *Radiology* **228**, 23–25 (2003)
 40. Westergaard, D., Stærfeldt, H.H., Tønsberg, C., Jensen, L.J., Brunak, S.: A comprehensive and quantitative comparison of text-mining in 15 million full-text articles versus their corresponding abstracts. *PLoS computational biology* **14**(2), e1005962 (2018)