DOI: 10.1561/9781638282372.ch19

Chapter 19

Implementation and Impact of AI for the Interpretation of Lung Diseases in Chest CTs

By Sebastian Roehrich, Helmut Prosch, Mia Ilic, Allan Hanbury, Georg Langs and Markus Krenn

19.1 Introduction

This chapter describes the implementation and integration of an Artificial Intelligence (AI)-based Content-Based Image Retrieval (CBIR) system in a clinical setting at the Medical University of Vienna and the Vienna General Hospital between 2019 and 2020. The system enables radiologists to find similar cases during the assessment of lung diseases in chest Computed Tomography (CT) data and was implemented and validated in a radiology department in the course of the BigMedilytics (BML) project.

19.2 Clinical Need and Context

In daily clinical practice, radiologists face a rapidly increasing volume of data or cases. They need to assess more imaging data and, at the same time, deal with a growing complexity of diagnoses and corresponding treatments. This causes a gap between the number of available experts and their capacity. AI solutions that can reduce the time needed to reach a diagnosis and to improve diagnostic accuracy are therefore highly relevant.

On average, radiologists see structures (i.e., pathology) with which they are not familiar in some 20% of CT or Magnetic Resonance (MR) images. While this percentage varies with experience and width of knowledge, it constitutes a substantial part of studies viewed on a daily basis. The process of finding additional information in these cases, in order to write a report, takes up to ±20 minutes (e.g., involving asking colleagues, paging through reference books, performing online searches, or consulting other sources).

When assessed by multiple radiologists, there is a significant variation in the identification of pathologies in the same images, for example, in pneumonia [1]. This suggests that pinpointing a diagnosis is challenging given the information typically available. Errors and discrepancies in practice are uncomfortably common, with an estimated day-to-day rate of 3–5% of studies reported, while even higher rates have been reported in publications [2].

The increasing gap between the volume of medical imaging data and the number of qualified radiologists makes the increase in efficiency a pressing issue. The improvement of quality by leveraging knowledge encoded in more than 1 billion unused CT and MRI scans in Europe is critical to provide fast and high-quality diagnostics to the European population. Prototype search software was developed to tackle this problem by (1) enabling fast and effective access to, and use of large medical imaging databases and (2) enabling clinicians to deliver higher accuracy diagnoses in a smaller amount of time, with the aim that the outcome will have a direct impact on the clinical productivity of radiologists and medical professionals using imaging data.

In the course of the BML project, the contextflow integrated prototype software was used to improve the radiology workflow in a clinical setting. It aimed at reducing the time to diagnosis in radiology departments and at the same time improving the quality of diagnosis by providing an efficient search engine for digitally available comparative radiological data. Using the prototype, radiologists could access comparable cases, connected information, and reference cases useful for a differential diagnosis, based on visual queries in the imaging data they were reading. The increase in diagnosis efficiency and the ability to effectively search in large databases of medical imaging data is critical as about 30% of worldwide storage capacity will be occupied by biomedical imaging data over the next years with more than 125 million CT and MR examinations being performed yearly in the EU alone [3].

Radiologists need rapid access to information and document evidence to back up their initial interpretation of the images before formulating a diagnosis. External resources are being used in about 20% of cases, consuming a significant amount of time. In many cases, radiologists need to ask their colleagues or search for reference literature or web resources, which is time-consuming and prone to errors.

Generally, radiologists follow the same procedure from opening a lung-CT case until the report is finished. As a first step, the predominant pattern in the image(s) needs to be identified. The software and CBIR may already help at this stage by providing an automated estimate of the quantity and location of different disease patterns. Next, the spatial distribution of the patterns needs to be assessed. By providing mapping in the CT image of the distribution of pathologies, a quick estimate of the distribution can be made at a single glance (instead of scrolling through the whole volume and using different views). Then, additional findings have to be taken into consideration to narrow the list of differential diagnoses. Again, the quantification and detection of 19 different patterns by contextflow may help at this stage of the diagnostic process.

Finally, all of the abovementioned imaging findings need to be put together to formulate a main diagnosis and, if necessary, several differential diagnoses. In order to support radiologists during this task, contextflow provides information relevant for interpreting image findings such as lists of relevant diagnoses, tips, and possible pitfalls together with references to external resources such as Radiopaedia or STATdx.

Two quality indicators in radiology workflow are reading time and diagnostic quality. To assess the impact of the prototype on these indicators, the final software version was deployed at the Department of Biomedical Imaging and Image-guided Therapy of the Medical University of Vienna, and a reader study to evaluate the software prototype was conducted at the Medical University of Vienna, with eight radiologists. The design was chosen to resemble clinical routine as much as possible in terms of case variety and the process of reading to further improve comparability and integration into the daily radiological workflow [4].

19.3 The AI Solution

The CBIR system that was implemented in the clinical setting as part of BML consisted of software that analyzes CT data and searches for similar cases based on marked regions of interest. The user assesses a patient's CT and marks a Region Of Interest (ROI) in the volume data. The software extracts features and searches for similar patterns in thousands of reference cases in less than a second. It presents the resulting cases together with suggested descriptions of the findings and relevant information for differential diagnosis (see Figure 19.1).

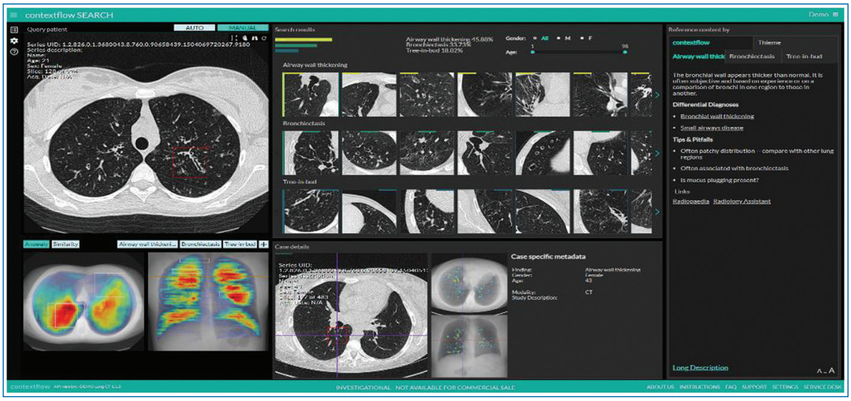

Figure 19.1. The contextflow search prototype, which was used during the intervention phase, is a web application executable from the local PACS. (1) The radiologist initiates the search for similar cases by drawing an ROI in the current CT scan. (2) A "heat-map" in the lower left corner visualizes and quantifies the distribution of one of 19 selectable lung patterns for the current scan. (3) Similar cases to the three most predominant patterns in the ROI are shown arranged according to the highest lung pattern classification probability. Choosing one case leads to (4) more information about the visually similar case (bottom middle). (5) Relevant content to the predominant pattern is presented as a list of differential diagnoses with links to the respective Radiopaedia.org page, tips and pitfalls for the patterns, and additional in-product content for differential diagnoses (right-hand side).

During a typical assessment of radiological imaging data, the radiologist takes into consideration reducing the images to their component part, reports on findings, and, in difficult cases, consults a range of sources, to identify the finding, verify suspected findings, or put the finding in the context of the disease. The software prototype developed for BML supported this by enabling radiologists to trigger searches by marking an ROI in the imaging data. The software then compared the marked patterns with a large database of cases, ranked cases, and showed the most similar cases, together with a summary and scoring of findings, and additional information such as differential diagnosis guidance, or direct links into curated literature optimized for supporting radiologists.

This contextflow search system allows search in radiology image archives containing image examples of a wide variety of diseases with accompanying radiology reports. The focus of the study was lung diseases. It allowed a radiologist to select an ROI in the 3D CT image. The deployment of the contextflow prototype consisted of the integration into the hospital infrastructure and the on-site Picture Archiving and Communication System (PACS) used for managing and viewing medical images. Integrating the contextflow search software directly into the PACS facilitates the integration of the search capability into existing workflows of radiologists easily accessible through an additional search button on their standard image viewing (PACS) interface. The system was deployed at the Medical University of Vienna.

In general, deployment is an important part of enabling effective and efficient use for a system deeply integrated into the daily routine of healthcare workers. In the project, deployments were managed by a single Deployment Master node that coordinated the deployment for all sites and had all deployable packages available. Two kinds of deployment sites were present—first, hosts which are globally accessible and, second, hosts which are located within the protected intranet of a hospital. The Retrieval Backend is generally deployed in the first scenario. The Application Backend is usually deployed inside a protected intranet, but public instances can be deployed as well.

19.3.1 User Interface for Radiologists

The Contextflow platform was used by radiologists reading individual patient cases. They used the platform via their viewer or a browser-based interface to obtain relevant information for the current case. The user interface enabled the radiologists to (1) trigger a search by marking an ROI in the image or volume in front of them and (2) view and explore the search results. The user interface enabled the user to:

- Trigger a search based on a marked ROI
- View search results and their statistical characteristics (e.g., findings)
- Group search results
- Explore search results by providing a detailed view in which the user can inspect all cases in the result list, scrolling through the volume, and visualizing the distribution of areas similar to the query
- Find reference information relevant to the diagnosis and differential diagnosis

19.3.2 Search Model and Engine

During a search, the radiologist inspects an image and marks an ROI to indicate a pattern that should serve as the basis for the search. This information is sent to the search model and engine as a query. Together with the query image, the retrieval engine can receive the image information together with other data such as location or patient-specific information. The search is performed on a local image content level. The index consists of billions of image locations across thousands of reference CT volumes. Triggered by the query ROI, locations with similar appearance are found and ranked across the index, and cases are then ranked corresponding to the found image regions they contain. The retrieval unit ranks the images based on the query and enables further filtering with query information. It provides the results either as an output to the user at the user interface or as an input to further processing. The retrieval unit can perform multiple searches serving different users simultaneously and can also search multiple indices.

19.4 AI Components

The contextflow search platform performs CBIR in the biomedical domain using deep-learning techniques such as Convolutional Neural Networks (CNNs) for data processing. To initiate a search, the user marks an ROI containing the pattern in the case that is being assessed. Based on this query, it ranks indexed examples corresponding to their visual similarity to the query imaging data. These ranked examples form the query result. When receiving the query, the platform can also pseudo-anonymize the query data in the browser.

The AI components solve two problems. First, they have to learn an effective visual similarity function that captures disease-relevant similarity, as a substantial amount of variability exists but is not linked to the diagnosis. Second, the component needs to accelerate the comparison of a query with billions of examples. As the search is performed on a region level, many thousands of regions are indexed for every image volume in the search database. In practice, this amounts to billions of entries, for which the similarity measure has to be evaluated in the time between the user query and the display of the search results. Machine learning algorithms are used to process imaging data, learn which features to extract, and how to compare them, and are active to conduct the search. Machine learning is a crucial part of image processing as appearance differences associated with disease and diagnosis are often subtle compared to the overall variability in the normal population. However, here they have to drive the image search during radiological diagnosis. Deep learning was used to train the models on lung diseases and to compare measures based on imaging data reflecting the disease-specific appearance.

19.4.1 Indexing

The indexing engine is given a dataset and optimized metrics that have been taught during a training phase. It creates a structure holding the data and facilitates finding data similar to a query case. It is optimized to store the information of image features, optionally together with metadata information, and to enable parallel searches. The indexing uses the trained machine learning model that quantifies a metric between image patches. In the resulting representation, simple distances can be used to rank cases reflecting the similarity of visual information. That is, after mapping to the embedding space by the learned representation function, a simple distance such as the Euclidean distance is used for the ranking of similar cases. This yields a representation of the entire dataset. For each representation of a region, it is also known from which volume at which position it comes from so that once a region is identified as a high-ranked match, a user can be presented with the corresponding image.

19.4.2 Retrieval

Given an image, and optionally a user-indicated ROI, retrieval finds the most similar image representations in the index and returns a sorted list of regions and corresponding images. These results can be extended by information (e.g., textual information, specific structured information, or variables) linked to the volumes contained in the list. The retrieval unit can also perform retrieval based on whole images or parts of the images. In either case, the ranking of retrieval results can be presented on the block or volume level.

The retrieval can be performed either as a single retrieval or as a Retrieval Cascade (Figure 19.2), where the result of one retrieval (e.g., a weighted list of textual terms associated with the top-ranked examples) can serve as an enrichment of a subsequent retrieval step that uses both the initial query and the enrichment as input. For example, the retrieval of similar cases can yield statistics about the terms in radiology reports associated with the top-ranked cases. In a second retrieval, the initial ROI and image information together with these candidate terms are used to perform a retrieval in other sources.

19.5 Integration Into Clinical Settings

For a seamless use of the contextflow search system prototype, the software was directly integrated into the local PACS of the Department of Biomedical Imaging and Image-Guided Therapy at the Medical University of Vienna and the General Hospital of Vienna (Austria). This integration allowed the evaluation to be under circumstances that resemble clinical routine. Looking at the early observations in the project study, we identified the need to close the gap between the routine use of the PACS and the image retrieval prototype to minimize the time needed for its use. Using the experience from this integration, the contextflow search system was also integrated into other PACS, enabling a so-called 'deep integration' into the radiology workflow; among them are Philips, Medigration, and Sectra. The framework for deployment and updating of the contextflow search system was created, allowing straightforward deployment and maintenance over installations in multiple hospitals.

The key goals of deep integration were as follows: (1) making the prototype available within the typical working environment and systems of the radiologists (2) testing its usefulness in terms of workflow and assessment support on site.

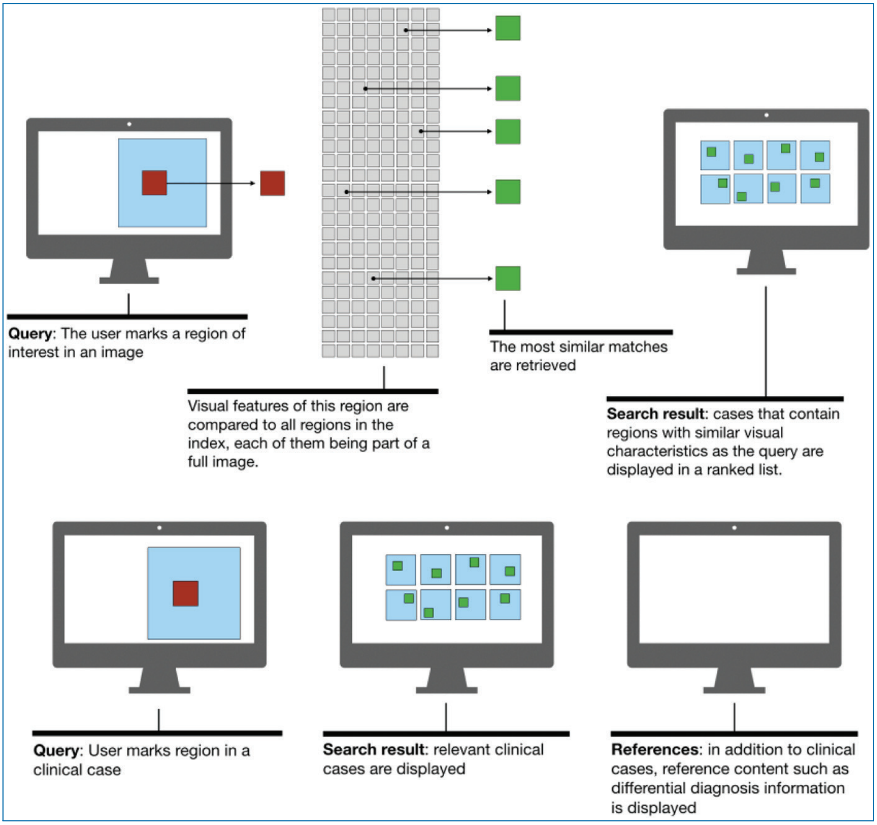

Figure 19.2. User query based on image content, after training and indexing are finished.

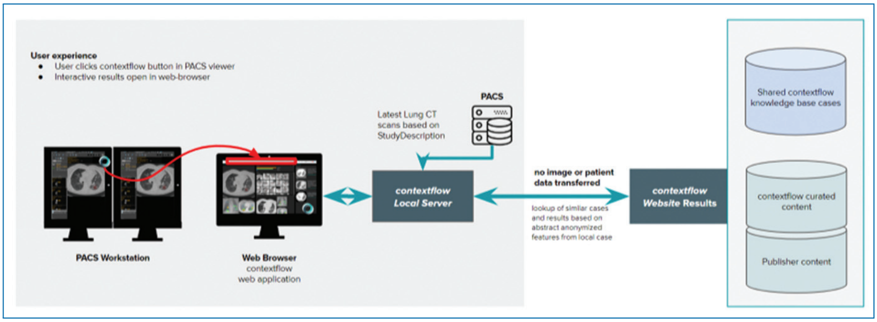

Figure 19.3. Integration of contextflow SEARCH into the PACS workflow.

A context diagram of the system integration is shown in Figure 19.3, illustrating the integration of the system into the clinical routine of the hospital.

A physician interacts with the system via the graphical user interface (application frontend) of the search application component. Here, they select a dataset to examine, execute queries, and explore search results. The system is integrated into the Information Technology (IT) infrastructure (e.g., in a hospital). The PACS system pushes datasets that can be examined into the system. Subsequently, these datasets get selectable by the user. Search results are further linked to additional information such as metadata and reports and other sources such as publications, articles, and curated knowledge bases. The system is implemented with three major components, each of which covers different system requirements:

- The search Application Frontend is the user interface that allows selecting datasets, executing queries, and retrieving and exploring search results.
- The Application Backend is responsible for preparing incoming datasets for examination, managing query execution, combining query results, and presenting this functionality in an API that is used by the Application Frontend.
- The Retrieval Backend performs retrieval queries and serves linked data such as images, meta information, and other sources.

19.6 Clinical Evaluation of the Software

While there is research on technical aspects of medical image processing, including the retrieval of semantic information, there is little research published on user evaluation of the technology and clinical utility. The technology can be useful for tackling diagnostic problems by providing radiologists with similar cases and additional information such as online reference content and thus improving the radiological workflow. In clinical settings, it means potential improvement in reporting on interstitial lung diseases or chronic obstructive pulmonary disease, for example [5].

The goal of the study was to measure the impact of the tool in the clinical settings on the radiologists' workflow of interpreting pulmonary chest CTs by allowing them access to additional, relevant information which they could use at their convenience.

Such evaluation is valuable for proving there is a connection between the CAD tool and an improvement in diagnosis or the workflow turnaround time, which would show the clinical benefit of the CAD tool.

19.6.1 Materials and Methods

For the reader study, the database of query cases held 108 chest CTs obtained from five scanner manufacturers in 2018, of which 100 cases had a confirmed diffuse parenchymal lung disease, and 8 cases without. Each CT had a diagnosis confirmed by a sub-specialized thoracic radiologist with 20 years of experience using

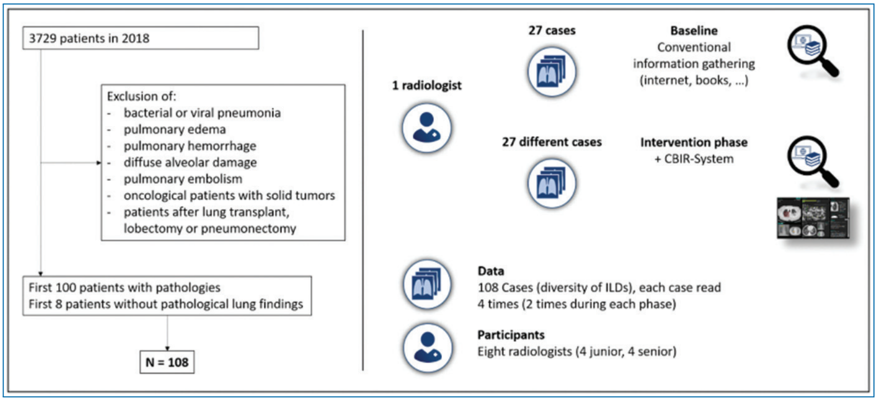

Figure 19.4. Left: Exclusion and inclusion criteria. Right: Distribution of cases [4].

the existing clinical information available in the patient records. Ethical approval was obtained as part of the BigMedylitixs project.

Chest CTs were blinded to eight radiologists, divided into four junior and for senior professionals. They were each given 54 pulmonary CT cases – chosen at random – to read and report in two phases in a setting that resembled the realistic clinical one: each radiologist read 27 CTs without support from the contextflow search prototype as part of the baseline phase and another 27 CTs with the contextflow search prototype as part of the intervention phase. There was a washout period between the reads. In the end, each of the 8 participants had read their 54 unique cases, resulting in a total of 432 readings of cases (Figure 19.4).

During both reads, the participants were allowed access to additional information of their choice such as books or online literature, but not to intercollegiate discussion. In the intervention phase, they also had access to the reference content of the contextflow search prototype. The cases selected were unknown to the participating radiologists, and a 'correct answer' for each was available for comparison purposes.

19.6.2 Study Findings

The system was evaluated in a multireader study with 8 radiologists creating 430 reports. The study showed that time savings of more than 30% were achieved by using the AI system as part of the case assessment. Results were published in 2022 [4].

The overall turnaround time of the radiological workflow per case decreased by 31.25% ($p < 0.001$) when the contextflow prototype was used (Figure 19.5). As the participants had access to reference content in both reads. Additional content was available in the intervention phase in the form of AI, and the results show they

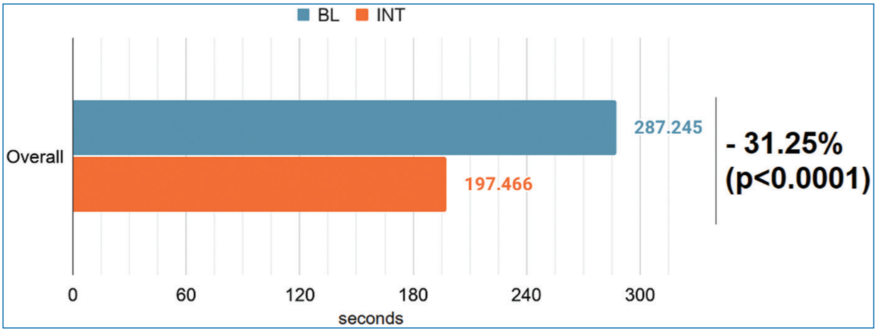

Figure 19.5. Overall average reporting time per chest CT case corrected for other factors such as seniority without the contextflow tool (baseline, BL) and with it (intervention, INT), in seconds [4].

searched more often during the intervention phase (24%) despite the reduction of time investment. Although the results show that the participants searched through more additional information while reading chest CTs, the participants needed less time to complete the cases when they searched for additional information in the contextflow tool (110 vs. 39 seconds used, $p = 0.002$).

Although not statistically significant, there is a tendency toward higher diagnostic correctness in the intervention read ($p = 0.083$).

19.6.3 Conclusion of the Reader Study

The results show that there is an effect of the CAD on the interpretation of chest CTs containing diffuse parenchymal lung disease and an improvement of the radiological workflow by lowering the reading time, in spite of the increased use of the relevant literature during the readings. Due to the use of the contextflow search system, the average time saved per read is 30%. This exceeds the increase in speed of 20% hypothesized at the beginning of the BML project.

It is important to look at this also from the disease diagnosis level. Although the evaluation does not show a significant improvement in diagnostic accuracy when a CAD is used, it shows there is no loss of diagnostic correctness and hence no negative impact on the diagnostics.

The prototype is based on lung patterns and retrieves visually similar cases without diagnosis information. However, in combination with a highly curated dataset, this technical construct has the potential for an upgrade, which would help diagnostic accuracy and correctness. The prototype developed during the project was used to develop a more comprehensive product that shows predominant patterns in chest CTs, gives quantification values for certain lung patterns, and retrieves visually similar cases from an internal database and is commercially available.

19.6.4 Limitations of the Reader Study

The participants were given chest CTs without accompanying clinical data or previous examinations. The study was interrupted by the COVID-19 outbreak, and the participants had a strongly variable washout period between 4 weeks and 15 months, and it may be expected that the participants, especially junior radiologists, gained more experience in the meantime.

19.7 Lessons Learned

This paragraph summarizes practical lessons learned during the implementation and evaluation of the system.

19.7.1 IT Requirements

Often the IT infrastructure in hospitals grows organically over the years, and it is challenging to implement substantial changes necessary for establishing novel technology. Innovations might need considerable effort to be deployed in this environment with complex infrastructure and the necessary framework of regulations of hospitals.

In cases where the hospital does not provide a sufficiently powerful computing environment or access is restricted, it may be necessary to install separate servers, sometimes with graphic processing units (GPUs), and integrate them into the existing IT infrastructure. This might increase the project (capital) expense, and the integration of new hardware might take time and will doubtlessly require approval from a number of departments. Overall, it is important to be flexible and be able to find quick solutions in a collaborative manner to make the system work. Effective and continual communication with the hospital IT departments is a key accelerator in this process.

Moving to a secure cloud infrastructure could help address these points and make innovation move faster. At the moment, the outlook and legislation (see Section I for extensive considerations) environment of hospitals in Europe is very different compared to the United States when it comes to cloud-based propositions. Innovation in healthcare and big data will most likely move faster in countries where the cloud is adopted earlier.

19.7.2 Imaging Data

Training AI models to analyze imaging data requires sufficient data of good quality and high-quality segmentation work by experts. Training AI requires big datasets with images from a large number of patients, from different CT scan vendors with different diseases and technical parameters. However, the imaging data available are scarce, and access to data is limited [6]. It is also tied to various regulatory and ethical approvals that need to be obtained in advance of the model training. This makes the process of data acquisition, preparation, and annotations lengthy and potentially very expensive. Additionally, imaging technology changes continually; thus, the training and adaptation of AI models is a critical and continuous process.

19.7.3 Clinical Validation

New AI solutions that may have an impact on the treatment selection have a direct impact on patients' health and must therefore be approved by regulatory bodies such as the European Medicine Agency (EMA) in Europe and United States Food and Drug Administration (FDA) in order to be sold and used on the market. However, regulatory approval is in some cases not enough and the solutions need further clinical validation to gain the trust of users. In addition, the explainability of AI needs to be addressed as buyers and users need to understand the software and the AI behind it in order to trust it and hence use it effectively.

References

[1] Walsh, Simon L. F., Lucio Calandriello, Nicola Sverzellati, Athol U. Wells, and David M. Hansell. 2016. "Interobserver Agreement for the ATS/ERS/JRS/ALAT Criteria for a UIP Pattern on CT." Thorax. https://doi.org/10.1136/thoraxjnl-2015-207252.

[2] Brady, Adrian P. 2017. "Error and Discrepancy in Radiology: Inevitable or Avoidable?" Insights into Imaging. https://doi.org/10.1007/s13244-016-0534-1.

[3] High Level Expert Group on Scientific Data. 2010. "Riding the Wave, How Europe Can Gain from the Rising Tide of Scientific Data, Final Report of the High Level Expert Group on Scientific Data." European Commission. https://ec.europa.eu/eurostat/cros/system/files/riding%20the%20wave.pdf.

[4] Röhrich, Sebastian, Benedikt H. Heidinger, Florian Prayer, Michael Weber, Markus Krenn, Rui Zhang, Julie Sufana, *et al.* 2022. "Impact of a Content-Based Image Retrieval System on the Interpretation of Chest CTs of Patients with Diffuse Parenchymal Lung Disease." European Radiology, July. https://doi.org/10.1007/s00330-022-08973-3.

[5] Pieler, M. *et al.* Evaluation of automatic volumetry of honeycombing and ground glass opacity patterns in lung CT scans (2022). European Congress of Radiology 2022. https://dx.doi.org/10.26044/ecr2022/C-15193.
[6] Hofmanninger, Johannes, Forian Prayer, Jeanny Pan, Sebastian Röhrich, Helmut Prosch, and Georg Langs. 2020. "Automatic Lung Segmentation in Routine Imaging Is Primarily a Data Diversity Problem, Not a Methodology Problem." European Radiology Experimental. https://doi.org/10.1186/s41747-020-00173-2.