A Multisite, Report-Based, Centralized Infrastructure for Feedback and Monitoring of Radiology AI/ML Development and Clinical Deployment

Menashe Benjamin1,* Guy Engelhard1, Alex Aisen2, Yinon Aradi1, Elad Benjamin2

1 Philips Algotec, a Philips Co.
2 Royal Philips
* Corresponding Author: menashe.benjamin@philips.com
{guy.engelhard, alex.aisen, yinon.aradi, elad.benjamin}@philips.com

ABSTRACT

An infrastructure for multisite, geographically-distributed creation and collection of diverse, high-quality, curated and labeled radiology image data is crucial for the successful automated development, deployment, monitoring and continuous improvement of Artificial Intelligence (AI)/Machine Learning (ML) solutions in the real world. An interactive radiology reporting approach that integrates image viewing, dictation, natural language processing (NLP) and creation of hyperlinks between image findings and the report, provides localized labels during routine interpretation. These images and labels can be captured and centralized in a cloud-based system. This method provides a practical and efficient mechanism with which to monitor algorithm performance. It also supplies feedback for iterative development and quality improvement of new and existing algorithmic models. Both feedback and monitoring are achieved without burdening the radiologist. The method addresses proposed regulatory requirements for post-marketing surveillance and external data. Comprehensive multi-site data collection assists in reducing bias. Resource requirements are greatly reduced compared to dedicated retrospective expert labeling.

KEYWORDS

Artificial intelligence; Image data labeling; Post marketing surveillance; Cloud infrastructure for AI development; Feedback and monitoring of AI; Interactive multimedia reporting.

HIGHLIGHTS

- AI/ML algorithms require feedback and monitoring from the field, for initial development, ongoing improvement and surveillance.
- Traditional retrospective expert review is expensive to use for feedback and doesn't address monitoring.
- Interactive radiology reporting with image location hyperlinks gathered from multiple sites and centralized in the cloud provides data for feedback and monitoring.
1. INTRODUCTION

The successful deployment of Artificial Intelligence (AI) and Machine learning (ML) in radiology presents many challenges [1-3]. Development of effective algorithms requires enormous amounts of data for training, testing and validation [4-6]. The best data is comprised of volumes of imaging studies collected from diverse sites, modality vendors, protocols and patient populations, with findings labeled by experienced radiologists. Labels are most useful when they identify locations on images. The strength of a label depends on many factors that affect its quality [7]. Such data is difficult to obtain on a large scale [8, 9]. Even in sophisticated facilities, it may be difficult to define, obtain and manage, especially when more than one institution is involved [10].

Once an algorithm is developed, tested, approved and deployed, regulators and good clinical practice require an ongoing process, preferably automated, for monitoring algorithm performance, such as by assessing whether radiologists agree with its output.

Labels for ML data sets have traditionally been obtained by retrospective expert review [9, 11]. Considerable effort is required to curate multi-system, multi-institutional, multi-national image collections and to perform retrospective expert labeling [12]. There is a trade-off between the effort it takes to label data, and the quality of the labels. Weakly labeled data may undermine learning performance [13], though in some circumstances even weak labels, when available in large numbers, may be sufficient [14]. A mixture of trusted and noisy labels may be useful [15].

For algorithms to be generalizable for deployment in diverse settings, performance should be verified using appropriate external data, beyond the data sets on which they are initially developed [16, 17]. Many published descriptions of algorithms lack such external validation [18]. Algorithms may fail when implemented in novel environments or presented with new data [19]. Generalizability is addressed by use of appropriate training data accounting for many factors, including disease prevalence and severity, technical modality considerations such as image resolution, varying acquisition protocols, modality vendor [20-22], and avoidance of ethnic, gender, and socioeconomic biases in the training data [5, 23-25]. Sample size, choice of validation strategy and appropriate metrics for performance evaluation are important [26].

To facilitate acquisition of the necessary training data, and to perform ongoing algorithm monitoring, appropriate tools must be integrated into the clinical workflow, allowing for practical and painless collection of radiologist feedback [27]. Algorithms must be evaluated in the context of their intended use by health care professionals, even to the extent of assessing their effect on patient outcomes when practical [17]. Yet large scale clinical trials for such evaluations may not be feasible or affordable, suggesting the need for pragmatic approaches such as assessment of performance in the field, recognizing that observational data is less robust than randomized data.

Current approaches, including retrospective expert annotation of limited datasets, are not integrated into the routine clinical workflow. They fail to provide a scalable mechanism for obtaining and recording feedback from radiologists. They also cannot monitor deployed...
algorithms to detect and recover from degraded performance and automate quality improvement.

To address these limitations, an infrastructure is needed for geographically-distributed creation, collection, and use of high-quality curated and labeled image data for developing and improving AI algorithms. Since radiologists will not use tools that reduce their productivity, an efficient approach to providing the necessary labels is to take advantage of their findings while studies are being interpreted, using reporting software that is seamless to their workflow.

1.1. The Cyclic Nature of AI Development: Feedback and Monitoring

AI/ML algorithm development is a cyclic process, in which algorithms can be developed, tested and improved in an iterative manner, and if sufficiently mature, redeployed in production, either continuously or at intervals (Fig. 1). We distinguish two steps in the process: Feedback and Monitoring. Ideally, both would be achieved by a method seamlessly integrated into the clinical workflow [28].

Fig. 1. The AI Development and Deployment Process – a continuous, never ending cycle. Until regulatory approval, algorithms are developed, tested and improved iteratively in the background using a feedback process. After regulatory approval, they are monitored for accuracy during routine reading against labels collected from interactive radiology reports.

Feedback [29] is the gathering of information about algorithm performance, including the output of the algorithm as run, and the truth with which to compare it. Such feedback may be the findings of an expert, a radiologist for instance, obtained prospectively (perhaps during clinical operation) or retrospectively. It may be collected with or without exposing the expert to
the algorithm output. Feedback based on reliable evidence of truth can be used for the labeling of data for training and testing of algorithms. It can be obtained in distinct testing phases, or in an operational deployment.

**Monitoring** is the continuous gathering of feedback during operation to determine if a deployed algorithm is performing as expected, e.g., consistently with the radiologist. Safety and efficacy can be continuously assured by monitoring. Monitoring will expose deterioration in performance, including internal drift (changes in model behavior) and external drift (changes in data characteristics) [30]. When deviance is detected, alerts can be raised, and provided to the user, management and the algorithm developer. An unexpected influx of patients with new characteristics, the presence of new diseases, or the deployment of new imaging equipment or techniques, can all lead to performance degradation, which in severe cases may be catastrophic. Monitoring, tightly integrated with routine clinical reporting, may be able to detect this earlier than other approaches. Detection of performance degradation at one site may serve to alert other sites. Monitoring also safeguards against malicious attacks.

Regulatory approval of AI/ML devices, whether they be locked successive releases, or continuously learning systems, requires greater emphasis on post-marketing surveillance for continued assurance of safety and efficacy, a so-called Total Product Lifecycle regulatory approach [31, 32]. Monitoring and feedback support surveillance. Feedback and monitoring information derived from clinical reports may satisfy regulators’ requirements for Real-World Data [33].

The cyclic process of improvement based on feedback must be rigorously documented, as a best practice and for regulatory compliance. Improvements must be reproducible, demonstrated and quantifiable, and serve to enhance safety and the AI value proposition. Continuous and significant improvement, as well as monitoring for performance degradation, may also serve to mitigate the phenomenon of AI "disillusionment" with underperforming algorithms [34], as well as engage the radiologists in their recognized role as information specialists [35]. Monitoring may mitigate issues similar to those evident with the early deployment of mammography Computer Aided Detection (CAD), wherein some studies suggested real world benefits did not meet initial expectations [15, 36]. The difficulty of gathering data for feedback and monitoring has been emphasized [37].

2. **APPROACH**

2.1. **Leveraging the Radiology Report**

The radiology report is a rich source of information [38] and may be a useful source of labels [9]. Retrospective natural language processing (NLP) can be applied to the routine clinical radiology report to obtain labels to train deep learning–based image classification models [39]. Such labels are limited and considered inferior to labels generated by explicit expert review and/or follow-up, but are easier and less expensive to obtain [40]. Concerns have been raised about the reliability of report-extracted labels, emphasizing the importance of the knowledge
extraction algorithm and choice of labels [41, 42]. Quantitative approaches to measuring the value of labels have been proposed [43].

Radiology reporting solutions may incorporate features such as the inclusion of multimedia, for example key or annotated images, and interactive functionality such as hyperlinks between report text and specific findings on images (Fig. 2).

We propose that NLP integrated into the reporting process, when coupled with hyperlinks to images and locations on images, can produce stronger labels than NLP alone, in the sense that one can be more confident they represent ground truth. Such labels are approaching the gold standard of dedicated retrospective expert review, with reduced effort and less impact on efficiency.

Localization of findings within images is important. Hyperlinks are often routinely inserted during reporting for use in linking prior studies or facilitating communication to the referring practitioner, when a suitable reporting system is used [44]. These have the significant benefit of connecting findings described in report text with images. Their use often requires little additional effort. In contrast, in the retrospective expert review scenario, precise localization may be "far too time-consuming, inaccurate, and not reproducible" [12]. However, localization information need not always be precise to be useful. Approximate information as from pointers or bounding boxes [45], typical of what radiologists use in routine clinical work, may be sufficient. The feasibility of extracting bounding boxes for AI/ML from annotations made routinely in the workflow has been demonstrated [46].

Capturing hyperlinks does not significantly disrupt the process of dictating a report, or distract the radiologist from looking at images, since a voice command may be used to insert the hyperlink, a very minimal incremental burden (Fig. 2). Radiologists create hyperlinks and reuse them when reporting new studies, even in the absence of evidence of use by referring providers [47]. Hyperlinks enable the automated generation of a multimedia report with pictures of key images, which may provide a competitive advantage if preferred by referrers [44]. Multimedia reports are increasingly desired throughout the enterprise [48]. Even in settings requiring a significant investment of radiologist or preprocessing staff effort, such as to perform tumor quantification for treatment monitoring, efficiency can be achieved using dedicated reporting software. The localized measurement annotations are also re-used for AI/ML feedback and monitoring [49, 50].

Our proposed approach leverages hyperlinks produced in routine clinical reporting to extract information that can be reused as a source of feedback and data for monitoring. This is most efficient if the reporting system is tightly integrated with the PACS workstation.
Fig. 2. The Interactive Report as a means of creating reliable labeled data by radiologists during routine reporting (a similar approach can be used for reporting in other specialties).

A. The Interactive Report contains embedded hyperlinks to locations and measurements in images, enabling the efficient generation of labeled data.

B. Creating hyperlinks is straightforward and easily accomplished during routine interpretation. The mechanism displayed in the illustration employs a single voice command (“Hyperlink”) to associate the text to the aneurysm location and measurement.
2.2. Centralized Cloud Implementation

How can labels created at a site during the course of routine clinical radiology workflow at multiple sites using the interactive reporting approach be used for feedback for model updates and performance monitoring (post-marketing surveillance)?

We propose that the traditional approach of centralizing the image and label data, which has been well-established both for research and registry applications, as well as for clinical data sharing, is practical and appropriate. Though patient consent, study de-identification, ownership and business entity relationship issues are challenging to solve, they are well understood.

Incorporation of centralized image and label data acquisition for AI/ML feedback and performance monitoring into the routine production systems obviates the need for each site to build dedicated infrastructures [10, 51].

When PACS with integrated images and reporting shared across related sites is implemented (i.e., all sites have access to each other’s data for clinical purposes), the centralization of images and labels is straightforward. With common access to imaging studies and supporting data, and a consistent reporting interface for all sites, even a radiology group that provides professional services to unrelated facilities can provide centralized labels [52]. Centralized data collection for feedback and monitoring will require de-identification of images and labels. The requirements for de-identification are well documented [53] and routinely applied in a manner that preserves their utility for algorithm improvement [54-56].

It is appropriate to use a cloud-based infrastructure for data collection and aggregation [57, 58]. Each source site can de-identify the images and accompanying label data before transmission to the cloud. Such a big data platform will allow multicenter research collaboration and support very large numbers of curated imaging studies, though from routine clinically deployed systems, in a similar manner to systems that have previously been used for research and development collaborations [59].

The result is a network of interconnected systems, in which the integrated PACS and reporting system at each site acts as a source of both images and labels (Fig. 3). This allows for centralized performance monitoring, and feedback for updating algorithms. The updated algorithms are then redistributed to the sites, whether deployed on-premise or as a cloud service, continuously or with successively approved releases.

Important functions of the centralized cloud component will include not only gathering of the data and model retraining, but also appropriate reporting of performance to sites, managers, algorithm developers and regulators, as well as maintenance of an audit trail of changes to software including neural network model weights.
Fig. 3. A Centralized, Cloud Based Infrastructure for Feedback and Monitoring of AI/ML Development and Clinical Deployment.

1 Sites contributing training data and/or running approved AI algorithms send images and related data (e.g. orders), interactive reports and labels to the cloud.

2,3 Developed (or existing) algorithms are tested for accuracy by the feedback system, which compares labels from interactive reports to the output of the algorithm, without burdening the radiologist. When training algorithms, discrepancies and their associated data are used to iteratively improve the algorithms. For approved algorithms, the monitoring system warns relevant sites against unacceptable degradation.

5 Results generated by approved algorithms (possibly in the form of a complete radiology report) are automatically presented to radiologists with each relevant study during the reading workflow.

4,6 De-identification and Algorithm Execution and Result Distribution can be performed either locally, centrally, or both.
3. DISCUSSION

3.1. Value of the Approach

We believe that information readily captured from radiologists during routine clinical workflow is valuable and should not be squandered, even if retrospective labeling and localization becomes tractable or scalable. Until now, the report has been an underutilized source of re-usable data. The core gaps in many clinical radiology systems are the absence of structured data and the lack of linkage of images to the report. Both are needed to provide a source of well-characterized structured information to serve as a source of definitive labels. An interactive reporting system that combines image viewing, hyperlinking and report dictation with NLP [44] can close those gaps.

We describe a cloud-based infrastructure that collects images and labels from participating sites. It uses existing reporting methodologies of voice dictation, voice commands, NLP, and hyperlinks tightly integrated into the PACS reporting workflow. Hyperlinking causes little interruption to the routine radiologist dictation work, but results in a better report. The infrastructure allows algorithm developers to develop, tune and deploy their AI/ML models. Once deployed, the infrastructure continues to collect feedback and to monitor the algorithms’ performance against real radiologist reports using the same collection strategies.

The value of this approach lies in the establishment of an AI/ML quality improvement infrastructure, leveraging labels acquired in routine reporting across multiple sites. This greatly simplifies the task of satisfying the requirements of users and regulators for safe and efficacious products. It allows for automatic deployment and ongoing enhancement of algorithms and monitoring of their performance. It provides larger, more complete cohorts for development of new algorithms, which reside in a centralized cloud repository. Gathering a vast pool of routine radiology data may also enable identification of rare and subtle cases, which might not otherwise be detected, collected and shared.

Radiologists are skeptical about AI, and the knowledge that they are contributing data to improve the algorithms that they use routinely may also serve to empower them [8], especially if they see regular performance improvements. The importance of user engagement in the AI process is well recognized [60].

This approach is applicable to multiple specialties and any form of imaging, not just radiology, since other specialists provide feedback as part of their viewing, analysis and documentation of findings. Findings may be in the form of a dedicated report (such as in histopathology [61]), or part of the clinical record in the EHR (such as in dermatology or ophthalmology), either of which may incorporate interactive and multimedia features.

Leveraging the advantages of the cloud is important. Most currently deployed PACS have a significant on-premise component for image and database storage, though there is movement to image sharing over the Internet [62] and eventually purely cloud-based PACS [63]. At some
point, the centralization of data for AI/ML will change from "site to cloud" transfer to "cloud to cloud" transfer. Since most sites already require an off-site backup mechanism for business continuity and disaster recovery [64], a cloud-based mirror of the image and label data is often already implemented and could be leveraged [7, 65]

3.2. Legal, Ethical, Moral and Business Considerations

Our proposed approach involves strictly voluntary participation. The technical design requires neither the engagement of every radiologist nor every site. It is independent of considerations of legal, ethical and moral aspects of secondary re-use of data, and can be deployed consistently with different interpretations of such factors. We recognize it is vital for every contributor to balance the complex aspects of stewardship, ownership, privacy, commercial interest, exclusivity, open access and public good [66, 67]. The proposed solution adheres to ethical principles regarding selling data. A recent article suggested that it is reasonable for corporate entities to profit from AI algorithms developed from clinical data, though unethical for them to profit from the data themselves [67]. Any approach needs to be robust to inevitable changes in acceptable data practices over time [68]. Compliance with requirements for data sharing and privacy principles, including sharing across borders, is essential [69, 70].

We believe that these issues are independent of the technical solution proposed. Appropriate safeguards will permit optimizing the utility of, and maximizing the re-use of, routine clinically acquired data in a beneficial and ethically sound manner.

3.3. Technical Considerations

We propose the use of tightly coupled and well-integrated systems that combine reporting with labeling, and in which user experience is seamless. A single supplier can provide a unified solution, or separate image viewing and reporting systems from multiple suppliers can be integrated by customizations to achieve the necessary functionality.

There are existing standards for the encoding of image selection information (DICOM KOS) and location (DICOM SEG, RTSS, SR) and labels and measurements (DICOM SR), and their use is important for the record to be interoperable and shareable. There are as yet no standards for orchestrating the interaction between the viewing and reporting systems at the granularity required for strong labeling. Accordingly, a single system that combines both viewing and reporting, implementing the necessary elements of localizing abnormalities through hyperlinking or other mechanisms, and NLP [44], is well positioned to provide relatively strong labels without additional effort by the user and without extra cost.

The human-driven reporting process often involves consolidation of information from multiple sources, including pre-population with procedure information [71], machine or technologist acquired structured data, such as contrast [72], radiation dose [73] or measurements [74, 75], as well as assistance by decision support structured templates and guidelines [76, 77]. Such clinical decision support mechanisms may be leveraged for AI/ML result
incorporation during radiology reporting, and may enable comparison of the machine and radiologist findings [78]. An infrastructure that provides for recording and centralizing image findings in the form of a registry [78, 79], would also enable feedback and monitoring, but would likely not address the need for localization, unless the source structured data includes links to images and finding locations, and a means to access the images.

There are additional issues with report extracted labels that must be addressed when labeling is based on hyperlinks in reports. These include the common situation in which multiple findings are present in a study, and matching image-related hyperlinks with dictated report information. Hyperlinks provide an advantage in that the localized image information is directly linked to corresponding information in the body of the report. There is also the significance of the absence of placement of a hyperlink by the radiologist; this may imply normality, or simply indicate that a finding wasn't worth the effort of hyperlinking. Hyperlinks to normal findings as a means of illustration (perhaps in response to an explicit question in the request) may also be confounding. Hyperlinks may also be implicitly biased towards more important or more subtle findings (perhaps those that particularly need pointing out). Radiologists might also mark only representative lesions, not all of them.

In addition, there is variation between experts [80], and combining multiple opinions for the same case may improve the strength of a label [81]. It is unusual for more than one radiologist to read the same case, except during peer review or in a training environment. However, we expect that the amount of data gathered from much larger numbers of individual contributors reporting many more cases may compensate for any actual or postulated relative weakness compared to what a consensus of experts dedicated to the task might obtain.

Such details will need to be addressed with respect to the strength and quality of the labels obtained, perhaps with refinements in the reporting interface, or by radiologist training. Data gathered from initial deployments will permit the improvement of the viewer, the reporting system and data gathering platform, not just the algorithms to which it is applicable.

It is tempting to include an ever more expansive set of data elements in what is gathered from the sites, particularly to increase the strength of the labels and bring the performance monitoring closer to the actual patient outcomes. For example, the radiologist's clinical report does not contain the final diagnosis, which may reside in the EHR. Though HL7 feeds from the EHR can be incorporated in our proposed solution, there are risks in receiving a comprehensive unfiltered stream [82, 83]. Our initial focus on the radiology report and the image-associated content is limited, but it is aligned with a sufficient set of use cases to be adequate as a first step. More information from related systems can be incorporated in the automated data capture as the value of additional elements for model improvement is demonstrated.

3.4. Comparison with Alternative Approaches

The effort required for retrospective expert review outside of the clinical routine is considerable. It has been quantified for some tasks, and there is recognition that it may not be
sustainable beyond short bursts of activity and with limitations on the total effort per expert that is feasible (or affordable) [12]. Retrospective expert review also does not provide a solution for continuous performance monitoring.

Expert labeling and annotation platforms have been developed that seek to minimize the amount of effort required, and which optimize the capture of specific labels with or without localization [12, 29, 45, 84, 85]. Such systems long predate the specific use of labels for modern ML, but were applicable for mammography CAD research and teaching [86].

Dikici et al [29] describe a system that is dedicated to the purpose of acquiring feedback, but its use requires a custom-built tool and is outside the routine clinical workflow. They recognize that the use of DICOM standards allowing for interoperability of annotations mitigates this concern somewhat. They define stages of algorithm maturity in terms of distinct levels of research, production and feedback. Their level of feedback maturity can be achieved using existing integrated off-the-shelf interactive multimedia reporting tools. These may be scaled to a multisite level, without requiring every user or site to learn and use a specialized tool, or limiting the generalizability of the result.

Another alternative approach is crowdsourcing, rather than use of carefully selected subject matter experts. This may produce weaker labels but on a larger scale at lower cost, and is theoretically possible even for image classification and diagnostic tasks [87, 88]. Existing image data sets with a known diagnosis can be augmented with localization and measurement information by such methods [84, 85]. It remains to be seen whether a crowdsourcing approach can be used for regulated devices. Crowdsourcing may produce labels for training and testing, but does not address the need for feedback and monitoring.

Structured reports (consistently organized reports with a standard outline, itemized sections and language) have long been touted as advantageous for data extraction. Yet radiologists have proven recalcitrant to their adoption on a large scale, particularly if additional effort is required during authoring [89]. They are slowly gaining favor and have renewed potential due to the demand for AI labels, particularly when localization is not required [90]. Despite the existence of standards for encoding structured reports in an interchangeable form, and recognition of their utility for AI [91], vendor adoption is poor. Notwithstanding this reluctance, reporting systems that support structured reporting should also support image labeling for AI. Relatively strong labels could be captured without burdening the reporting radiologist, particularly if localization information is also recorded.

Without a change in radiologists' incentives, it seems that a more pragmatic approach than conventional structured data entry is required, one that leverages what can be captured through existing practice, to mitigate the risk of a "productivity nightmare" [89]. The use of disease-specific templates during conventional dictation with speech recognition coupled with NLP would facilitate structured data entry, as can the use of assistive technology during dictation [92]. Scalability of NLP for global use in a multitude of languages remains challenging.
Others have proposed the centralization of data with an "honest broker", as might be provided by professional societies [93], trade organizations, industrial consortia or other third parties, whether for profit or not. We suggest it is more practical to leverage an existing deployed PACS infrastructure, and make use of the system supplier to manage, host and make use of the centralized information. Our approach obviates the need to develop, interface, deploy and test additional tools at the source sites, which may be accompanied by non-trivial capital or operational costs. It also allows existing organizational and business relationships between site and provider to be leveraged and reused, without the need for cooperation with an additional external entity that may have a potentially different political or strategic agenda.

Various flavors of federated, distributed or "no peek" learning mechanisms have been proposed that leave the images and labels at each site and instead involve redistributing code or model weights [94, 95]. These have been demonstrated for radiology applications [96-98]. It remains to be seen how successful or practical such approaches will be on a large scale. Standardizing packaging of algorithms or model weights of a family of algorithms, for interchange between different platforms at different sites, is elusive.

The federated approaches require each site to use standardized labels and interchange formats, provide data warehousing and curation services and operate dedicated computational resources. This is challenging enough for major academic centers but likely to be impractical for small community hospitals and imaging centers from whom data is essential to achieve generalizability and avoid bias. Each site needs to deploy its own local feedback and monitoring solution if it is unwilling to centralize images and labels. A combination of federated learning approaches for training, and centralized feedback and monitoring, could coexist.

In federated schemes, the distribution of incentives such as credit and revenue may also pose business model challenges.

4. CONCLUSIONS

As the tools of artificial intelligence play an increasing role in the medical and business practice of radiology, consideration must be given to means of collecting and curating the data through which algorithms are trained. Mechanisms must be implemented to allow for the ongoing, site-specific evaluation of deployed algorithms; such surveillance will be required by regulators. Updated algorithms with enhanced performance can then be re-deployed.

We propose a cloud-based multisite infrastructure to accomplish these goals. We have defined its essential technical and operational characteristics. It includes the centralized collection of large amounts of de-identified image data from sites around the world, and the incorporation of tools to label the images seamlessly and accurately as the studies are routinely interpreted by radiologists.
The business model for AI/ML remains elusive at this nascent stage of development [99, 100]. We feel that any approach that can monitor and maximize quality, while at the same time minimizing effort and cost, should be explored.

REFERENCES

[1] K. Yasaka, O. Abe, Deep learning and artificial intelligence in radiology: Current applications and future directions, PLoS Med 15(11) (2018) e1002707-e1002707, https://doi.org/10.1371/journal.pmed.1002707.

[2] A. Hosny, C. Parmar, J. Quackenbush, L.H. Schwartz, H.J.W.L. Aerts, Artificial intelligence in radiology, Nat Rev Cancer 18(8) (2018) 500-510, https://doi.org/10.1038/s41568-018-0016-5.

[3] B. Allen, S.E. Seltzer, C.P. Langlotz, K.P. Dreyer, R.M. Summers, N. Petrick, D. Marinac-Dabic, M. Cruz, T.K. Alkasab, R.J. Hanisch, W.J. Nilsen, J. Burleson, K. Lyman, K. Kandarpa, A roadmap for translational research on artificial intelligence in medical imaging: From the 2018 National Institutes Of Health/RSNA/ACR/The Academy Workshop, JACR 16(9) (2019) 1179-1189, https://doi.org/10.1016/j.jacr.2019.04.014.

[4] J. Cho, K. Lee, E. Shin, G. Choy, S. Do, How much data is needed to train a medical image deep learning system to achieve necessary high accuracy?, arXiv preprint arXiv: 1511.06348, 2015, (Accessed on July 17, 2020).

[5] G. Choy, O. Khalilzadeh, M. Michalski, S. Do, A.E. Samir, O.S. Pianykh, J.R. Geis, P.V. Pandharipande, J.A. Brink, K.J. Dreyer, Current applications and future impact of machine learning in radiology, Radiology 288(2) (2018) 318-328, https://doi.org/10.1148/radiol.2018171820.

[6] J.A. Dunnmon, D. Yi, C.P. Langlotz, C. Ré, D.L. Rubin, M.P. Lungren, Assessment of convolutional neural networks for automated classification of chest radiographs, Radiology 290(2) (2019) 537-544, https://doi.org/10.1148/radiol.2018181422.

[7] M.J. Willemink, W.A. Koszek, C. Hardell, J. Wu, D. Fleischmann, H. Harvey, L.R. Folio, R.M. Summers, D.L. Rubin, M.P. Lungren, Preparing medical imaging data for machine learning, Radiology 295(1) (2020) 4-15, https://doi.org/10.1148/radiol.2020192224.

[8] B. Allen, S. Agarwal, J. Kalpathy-Cramer, K. Dreyer, Democratizing AI, JACR 16(7) (2019) 961-963, https://doi.org/10.1016/j.jacr.2019.04.023.

[9] C.P. Langlotz, B. Allen, B.J. Erickson, J. Kalpathy-Cramer, K. Bigelow, T.S. Cook, A.E. Flanders, M.P. Lungren, D.S. Mendelson, J.D. Rudie, G. Wang, K. Kandarpa, A roadmap for foundational research on artificial intelligence in medical imaging: From the 2018 NIH/RSNA/ACR/The Academy Workshop, Radiology 291(3) (2019) 781-791, https://doi.org/10.1148/radiol.2019190613.

[10] K. Magudia, C.P. Bridge, M. Walters, A. McCarthy, M. Michalski, K.P. Andriolet, M.H. Rosenthal, The trials and tribulations of assembling large datasets for machine learning applications, SIIM, Denver, CO, https://cdn.ymaws.com/siim.org/resource/resmgr/siim2019/posters/The_Trails_and_Tribulations_.pdf, (Accessed on July 17, 2020).

[11] E. Montagnon, M. Cerny, A. Cadrin-Chênevert, V. Hamilton, T. Derennes, A. Ilinca, F. Vandebroucke-Menu, S. Turcotte, S. Kadoury, A. Tang, Deep learning workflow in radiology: A primer, Insights Imaging 11(1) (2020) 22-22, https://doi.org/10.1186/s13244-019-0832-5.
A Multisite, Report-Based, Centralized Infrastructure for Feedback and Monitoring of Radiology AI/ML Development and Clinical Deployment

[12] A.E. Flanders, L.M. Prevedello, G. Shih, S.S. Halabi, J. Kalpathy-Cramer, R. Ball, J.T. Mongan, A. Stein, F.C. Kitamura, M.P. Lungren, G. Choudhary, L. Cala, L. Coelho, M. Mogensen, F. Morón, E. Miller, I. Ikuta, V. Zohrabin, O. McDonnell, C. Lincoln, L. Shah, D. Joyner, A. Agarwal, R.K. Lee, J. Nath, Construction of a machine learning dataset through collaboration: The RSNA 2019 brain CT hemorrhage challenge, Radiology: Artificial Intelligence 2(3) (2020) e190211, https://doi.org/10.1148/ryai.2020190211.

[13] T. Wei, L.-Z. Guo, Y.-F. Li, W. Gao, Learning safe multi-label prediction for weakly labeled data, Machine Learning 107(4) (2017) 703-725, https://doi.org/10.1007/s10994-017-5675-z.

[14] C. Zheng, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, W. Liu, X. Wang, Deep learning-based detection for covid-19 from chest CT using weak label, 2020, http://www.medrxiv.org/content/10.1101/2020.03.12.20027185v2, (Accessed on August 18, 2020).

[15] H.-P. Chan, L.M. Hadjiiski, R.K. Samala, Computer-aided diagnosis in the era of deep learning, Med Phys 47(5) (2020) e218-e227, https://doi.org/10.1002/mp.13764.

[16] S.H. Park, K. Han, Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction, Radiology 286(3) (2018) 800-809, https://doi.org/10.1148/radiol.2017171920.

[17] X. Liu, L. Faes, A.U. Kale, S.K. Wagner, D.J. Fu, A. Bruynseels, T. Mahendiran, G. Moraes, M. Shamdas, C. Kern, J.R. Ledsam, M.K. Schmid, K. Balaskas, E.J. Topol, L.M. Bachmann, P.A. Keane, A.K. Denniston, A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: A systematic review and meta-analysis, The Lancet Digital Health 1(6) (2019) e271-e297, https://doi.org/10.1016/s2589-7500(19)30123-2.

[18] D.W. Kim, H.Y. Jang, K.W. Kim, Y. Shin, S.H. Park, Design characteristics of studies reporting the performance of artificial intelligence algorithms for diagnostic analysis of medical images: Results from recently published papers, Korean J Radiol 20(3) (2019) 405-410, https://doi.org/10.3348/kjr.2019.0025.

[19] J.R. Zech, M.A. Badgeley, M. Liu, A.B. Costa, J.J. Titano, E.K. Oermann, Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study, PLoS Med 15(11) (2018) e1002683-e1002683, https://doi.org/10.1371/journal.pmed.1002683.

[20] C.F. Sabottke, B.M. Spieler, The effect of image resolution on deep learning in radiography, Radiology: Artificial Intelligence 2(1) (2020) e190015, https://doi.org/10.1148/ryai.2019190015.

[21] P. Lakhani, The importance of image resolution in building deep learning models for medical imaging, Radiology: Artificial Intelligence 2(1) (2020) e190177, https://doi.org/10.1148/ryai.2019190177.

[22] J.E. van Timmeren, R.T.H. Leijenaar, W. van Elmpt, J. Wang, Z. Zhang, A. Dekker, P. Lambin, Test-retest data for radiomics feature stability analysis: Generalizable or study-specific?, Tomography 2(4) (2016) 361-365, https://doi.org/10.18383/j.tom.2016.00208.

[23] M.A. Gianfrancesco, S. Tamang, J. Yazdany, G. Schmajuk, Potential biases in machine learning algorithms using electronic health record data, JAMA Intern Med 178(11) (2018) 1544-1547, https://doi.org/10.1001/jamainternmed.2018.3763.

[24] M. Goyal, T. Knackstedt, S. Yan, S. Hassanpour, Artificial intelligence-based image classification for diagnosis of skin cancer: Challenges and opportunities, arXiv preprint arXiv: 1911.11872, 2019, (Accessed on July 17, 2020).

[25] B. Allen, K. Dreyer, The role of the ACR Data Science Institute in advancing health equity in radiology, JACR 16(4) (2019) 644-648, https://doi.org/10.1016/j.jacr.2018.12.038.
[26] G.S. Handelman, H.K. Kok, R.V. Chandra, A.H. Razavi, S. Huang, M. Brooks, M.J. Lee, H. Asadi, Peering into the black box of artificial intelligence: Evaluation metrics of machine learning methods, AJR 212(1) (2019) 38-43, https://doi.org/10.2214/ajr.18.20224.

[27] T.S. Cook, The importance of imaging informatics and informaticists in the implementation of AI, Academic Radiology 27(1) (2020) 113-116, https://doi.org/10.1016/j.acra.2019.10.002.

[28] K. Juluru, H.-H. Shih, P. Elnajjar, A. El-Rowmeim, J. Fox, E. Siegel, K.N. Keshava Murthy, Building blocks for integrating image analysis algorithms into a clinical workflow, 2020, http://www.medrxiv.org/content/10.1101/2020.03.06.20027367v1, (Accessed on July 17, 2020).

[29] E. Dikici, M. Bigelow, L.M. Prevedello, R.D. White, B.S. Erdal, Integrating AI into radiology workflow: Levels of research, production, and feedback maturity, J Med Imaging (Bellingham) 7(1) (2020) 016502-016502, https://doi.org/10.1117/1.JMI.7.1.016502.

[30] G. Mahadevaiah, R. Prasad, I. Bermejo, D. Jaffray, A. Dekker, L. Wee, Artificial intelligence-based clinical decision support in modern medical physics: Selection, acceptance, commissioning, and quality assurance, Med Phys 47(5) (2020) e228-e235, https://doi.org/10.1002/mp.13562.

[31] B. Sahiner, B. Friedman, C. Linville, C. Ipach, E. Montgomery, Perspectives and best practices for artificial intelligence and continuously learning systems in healthcare, https://www.ehidc.org/sites/default/files/resources/files/Perspectives%20and%20Best%20Practices%20for%20Artificial%20Intelligence%20and%20Continuously%20Learning%20Systems%20in%20Healthcare.pdf, (Accessed on July 17, 2020).

[32] US FDA. Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD) - discussion paper and request for feedback., http://www.fda.gov/media/122535/download, (Accessed on July 17, 2020).

[33] Use of real-world evidence to support regulatory decision-making for medical devices, https://www.fda.gov/regulatory-information/search-fda-guidance-documents/use-real-world-evidence-support-regulatory-decision-making-medical-devices, (Accessed on July 17, 2020).

[34] H. Harvey, S. Islam, Blog: RSNA 2019 AI round-up. Hardian Health, http://hardianhealth.com/blog/rsna19, (Accessed on July 17, 2020).

[35] S. Jha, E.J. Topol, Adapting to artificial intelligence, JAMA 316(22) (2016) 2353, https://doi.org/10.1001/jama.2016.17438.

[36] C.I. Lee, N. Houssami, J.G. Elmore, D.S.M. Buist, Pathways to breast cancer screening artificial intelligence algorithm validation, Breast 52 (2020) 146-149, https://doi.org/10.1016/j.breast.2019.09.005.

[37] D. Higgins, V.I. Madai, From bit to bedside: A practical framework for artificial intelligence product development in healthcare, arXiv preprint arXiv: 2003.10303, 2020, (Accessed on July 17, 2020).

[38] J.M. Steinkamp, C. Chambers, D. Lalevic, H.M. Zafar, T.S. Cook, Toward complete structured information extraction from radiology reports using machine learning, J Digit Imaging 32(4) (2019) 554-564, https://doi.org/10.1007/s10278-019-00234-y.

[39] J. Zech, M. Pain, J. Titano, M. Badgeley, J. Schefflein, A. Su, A. Costa, J. Bederson, J. Lehar, E.K. Oermann, Natural language–based machine learning models for the annotation of clinical radiology reports, Radiology 287(2) (2018) 570-580, https://doi.org/10.1148/radiol.2018171093.

[40] J.J. Titano, M. Badgeley, J. Schefflein, M. Pain, A. Su, M. Cai, N. Swinburne, J. Zech, J. Kim, J. Bederson, J. Mocco, B. Drayer, J. Lehar, S. Cho, A. Costa, E.K. Oermann, Automated deep-neural-network
surveillance of cranial images for acute neurologic events, Nature Medicine 24(9) (2018) 1337-1341, https://doi.org/10.1038/s41591-018-0147-y.

[41] L. Oakden-Rayner, Exploring the Chestxray14 dataset: Problems, http://lukeoakdenrayner.wordpress.com/2017/12/18/the-chestxray14-dataset-problems/, (Accessed on July 17, 2020).

[42] L. Oakden-Rayner, Exploring large-scale public medical image datasets, Academic Radiology 27(1) (2020) 106-112, https://doi.org/10.1016/j.acra.2019.10.006.

[43] A. Ghorbani, J. Zou, Data shapley: Equitable valuation of data for machine learning, arXiv preprint arXiv: 1904.02868, 2019, (Accessed on July 17, 2020).

[44] S.D. Beesley, C.M. Gaskin, Interactive multimedia reporting: Key features and experience in clinical practice, JACR 15(8) (2018) 1120-1124, https://doi.org/10.1016/j.jacr.2018.03.027.

[45] G. Shih, C.C. Wu, S.S. Halabi, M.D. Kohli, L.M. Prevedello, T.S. Cook, A. Sharma, J.K. Amorosa, V. Arteaga, M. Galperin-Aizenberg, R.R. Gill, M.C.B. Godoy, S. Hobbs, J. Jeudy, A. Laroia, P.N. Shah, D. Vummidi, K. Yaddanapudi, A. Stein, Augmenting the National Institutes Of Health chest radiograph dataset with expert annotations of possible pneumonia, Radiology: Artificial Intelligence 1(1) (2019) e180041, https://doi.org/10.1148/ryai.2019180041.

[46] H.M. Do, F. Farhadi, D. Jin, L.B. Machado, O. Marques, J. Yao, G. Cohen, Z. Xu, L.R. Folio, Optimal radiologist annotations and extraction for accelerating deep learning training, SIIM 2019, Denver, CO, https://cdn.ymaws.com/siim.org/resource/resmgr/siim2019/posters/Optimal_Radiologist_Do.pdf, (Accessed on July 17, 2020).

[47] S.D. Beesley, J.T. Patrie, C.M. Gaskin, Radiologist adoption of interactive multimedia reporting technology, JACR 16(4) (2019) 465-471, https://doi.org/10.1016/j.jacr.2018.10.009.

[48] L. Folio, K. Garriott, The role of interactive multimedia content in enterprise imaging, http://www.youtube.com/watch?v=6E06b_ewAYk, (Accessed on July 17, 2020).

[49] H.M. Do, L.G. Spear, M. Nikpanah, S.M. Mirmomen, L.B. Machado, A.P. Toscano, B. Turkbey, M.H. Bagheri, J.L. Gulley, L.R. Folio, Augmented radiologist workflow improves report value and saves time: A potential model for implementation of artificial intelligence, Academic Radiology 27(1) (2020) 96-105, https://doi.org/10.1016/j.jacr.2019.09.014.

[50] N. Goyal, A.B. Apolo, E.D. Berman, M.H. Bagheri, J.E. Levine, J.W. Glod, R.N. Kaplan, L.B. Machado, L.R. Folio, ENABLE (Exportable Notation And Bookmark List Engine): An interface to manage tumor measurement data from PACS to cancer databases, J Digit Imaging 30(3) (2017) 275-286, https://doi.org/10.1007/s10278-016-9938-1.

[51] G.A. Cohen, L.R. Folio, E. Jones, J. Plum;, R. Levin, J. Galvez, H. Do, Big data management and archiving service for biomedical research AI/ML infrastructure, SIIM 2019, Denver, CO, https://cdn.ymaws.com/siim.org/resource/resmgr/siim2019/posters/Big_Data_Management_Cohen.pdf, (Accessed on July 17, 2020).

[52] M. Benjamin, Y. Aradi, R. Shreiber, From shared data to sharing workflow: Merging PACS and teleradiology, European Journal of Radiology 73(1) (2010) 3-9, https://doi.org/10.1016/j.ejradi.2009.10.014.

[53] National Electrical Manufacturers Association (NEMA). Digital Imaging And Communications In Medicine (DICOM) Standard PS3.15 - Security and System Management Profiles - E.1 Attribute Confidentiality Profiles - De-Identifier,
A Multisite, Report-Based, Centralized Infrastructure for Feedback and Monitoring of Radiology AI/ML Development and Clinical Deployment

http://dicom.nema.org/medical/dicom/current/output/chtml/part15/chapter_E.html#sect_E.1.1, (Accessed on July 17, 2020).

[54] J.B. Freymann, J.S. Kirby, J.H. Perry, D.A. Clunie, C.C. Jaffe, Image data sharing for biomedical research--meeting HIPAA requirements for de-identification, J Digit Imaging 25(1) (2012) 14-24, https://doi.org/10.1007/s10278-011-9422-x.

[55] S.M. Moore, D.R. Maffitt, K.E. Smith, J.S. Kirby, K.W. Clark, J.B. Freymann, B.A. Vendt, L.R. Tarbox, F.W. Prior, De-identification of medical images with retention of scientific research value, Radiographics 35(3) (2015) 727-735, https://doi.org/10.1148/rg.2015140244.

[56] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, F. Prior, The cancer imaging archive (TCIA): Maintaining and operating a public information repository, J Digit Imaging 26(6) (2013) 1045-1057, https://doi.org/10.1007/s10278-013-9622-7.

[57] G.C. Kagadis, C. Kloukinas, K. Moore, J. Philbin, P. Papadimitroulas, C. Alexakos, P.G. Nagy, D. Visvikis, W.R. Hendee, Cloud computing in medical imaging, Med Phys 40(7) (2013) 070901, https://doi.org/10.1118/1.4811272.

[58] L. Dai, X. Gao, Y. Guo, J. Xiao, Z. Zhang, Bioinformatics clouds for big data manipulation, Biol Direct 7 (2012) 43-43, https://doi.org/10.1186/1745-6150-7-43.

[59] E.L. Ridley, Israeli start-up eyes big-data tools for imaging analysis, AuntMinnie.com, http://www.auntminnie.com/index.aspx?sec=sup&sub=aic&pag=dis&ItemID=110966, (Accessed on July 17, 2020).

[60] R.W. Filice, R.M. Ratwani, The case for user-centered artificial intelligence in radiology, Radiology: Artificial Intelligence 2(3) (2020) e190095, https://doi.org/10.1148/ryai.2020190095.

[61] R. Colling, H. Pitman, K. Oien, N. Rajpoot, P. Macklin, V. Bachtiar, R. Booth, A. Bryant, J. Bull, J. Bury, F. Carragher, R. Colling, G. Collins, C. Craig, M.F. da Silva, D. Gosling, J. Jacobs, L. Kajland - Wilén, J. Karling, D. Lawler, S. Lee, P. Macklin, K. Miller, G. Mozolowski, R. Nicholson, D. O'Connor, M. Rahbek, N. Rajpoot, A. Sumner, D. Vossen, K. White, C. Wing, C. Wright, D. Snead, T. Sackville, C. Verrill, Artificial intelligence in digital pathology: A roadmap to routine use in clinical practice, J Pathol 249(2) (2019) 143-150, https://doi.org/10.1002/path.5310.

[62] M.B. Morgan, E. Young, S. Harada, N. Winkler, J. Riegert, T. Jones, N. Hu, M. Stein, Ditching the disc: The effects of cloud-based image sharing on department efficiency and report turnaround times in mammography, JACR 14(12) (2017) 1560-1565, https://doi.org/10.1016/j.jacr.2017.08.024.

[63] B. Siwicki, Cloud-based pacs system cuts imaging costs by half for rural hospital, Healthcare IT News, http://www.healthcareitnews.com/news/cloud-based-pacs-system-cuts-imaging-costs-half-rural-hospital, (Accessed on July 17, 2020).

[64] O.M. Ratib, B.J. Liu, J.M. McCoy, D.R. Enzmann, Application service provider (ASP) financial models for off-site PACS archiving, Proc SPIE Medical Imaging 2003: PACS and Integrated Medical Information Systems: Design and Evaluation, SPIE, 2003, pp. 57-62. http://dx.doi.org/10.1117/12.480464, (Accessed on August 18, 2020)

[65] P. Gerard, N. Kapadia, P.T. Chang, J. Acharya, M. Seiler, Z. Lefkovitz, Extended outlook: Description, utilization, and daily applications of cloud technology in radiology, AJR Am J Roentgenol 201(6) (2013) W809-11, https://doi.org/10.2214/AJR.12.9673.

[66] E.A. Krupinski, An ethics framework for clinical imaging data sharing and the greater good, Radiology 295(3) (2020) 683-684, https://doi.org/10.1148/radiol.2020200416.
[67] D.B. Larson, D.C. Magnus, M.P. Lungren, N.H. Shah, C.P. Langlotz, Ethics of using and sharing clinical imaging data for artificial intelligence: A proposed framework, Radiology 295(3) (2020) 675-682, https://doi.org/10.1148/radiol.2020192536.

[68] J.R. Geis, A. Brady, C.C. Wu, J. Spencer, E. Ranschaert, J.L. Jaremko, S.G. Langer, A.B. Kitts, J. Birch, W.F. Shields, R. van den Hoven van Genderen, E. Kotter, J.W. Gichoya, T.S. Cook, M.B. Morgan, A. Tang, N.M. Safdar, M. Kohli, Ethics of artificial intelligence in radiology: Summary of the joint European and North American multisociety statement, Insights Imaging 10(1) (2019) 101, https://doi.org/10.1186/s13244-019-0785-8.

[69] X. Larrucea, M. Moffie, S. Asaf, I. Santamaria, Towards a GDPR compliant way to secure European cross border healthcare industry 4.0, Computer Standards & Interfaces 69 (2020) 103408, https://doi.org/10.1016/j.csi.2019.103408.

[70] AMA privacy principles. 2020, 2020, http://www.ama-assn.org/system/files/2020-05/privacy-principles.pdf, (Accessed on July 17, 2020).

[71] C.M. Hawkins, S. Hall, J. Hardin, S. Salisbury, A.J. Towbin, Prepopulated radiology report templates: A prospective analysis of error rate and turnaround time, J Digit Imaging 25(4) (2012) 504-511, https://doi.org/10.1007/s10278-012-9455-9.

[72] S. Goldberg-Stein, D. Gutman, O. Kaplun, D. Wang, A. Negassa, M.H. Scheinfeld, Autopopulation of intravenous contrast type and dose in structured report templates decreases report addenda, JACR 14(5) (2017) 659-661, https://doi.org/10.1016/j.jacr.2016.10.017.

[73] M.D. Kovacs, M.Y. Cho, P.F. Burchett, M. Trambert, Benefits of integrated ris/pacs/reporting due to automatic population of templated reports, Curr Probl Diagn Radiol 48(1) (2019) 37-39, https://doi.org/10.1067/j.cpradiol.2017.12.002.

[74] N.J. Hangiandreou, S.F. Stekel, D.J. Tradup, Comprehensive clinical implementation of DICOM structured reporting across a radiology ultrasound practice: Lessons learned, JACR 14(2) (2017) 298-300, https://doi.org/10.1016/j.jacr.2016.09.046.

[75] M.H. Scheinfeld, O. Kaplun, N.A. Simmons, J. Sterman, S. Goldberg-Stein, Implementing a software solution across multiple ultrasound vendors to auto-fill reports with measurement values, Curr Probl Diagn Radiol 48(3) (2019) 216-219, https://doi.org/10.1067/j.cpradiol.2018.09.002.

[76] T.K. Alkasab, B.C. Bizzo, L.L. Berland, S. Nair, P.V. Pandharipande, H.B. Harvey, Creation of an open framework for point-of-care computer-assisted reporting and decision support tools for radiologists, JACR 14(9) (2017) 1184-1189, https://doi.org/10.1016/j.jacr.2017.04.031.

[77] M. Kohli, T. Alkasab, K. Wang, M.E. Heilbrun, A.E. Flanders, K. Dreyer, C.E. Kahn, Bending the artificial intelligence curve for radiology: Informatics tools from ACR and RSNA, JACR 16(10) (2019) 1464-1470, https://doi.org/10.1016/j.jacr.2019.06.009.

[78] M. Tilkkin, AI in radiology: Regulatory, quality, and implementation issues, http://on-demand.gputechconf.com/gtc/2018/presentation/S8994-workflow-and-regulatory-challenges-to-algorithm-implementation.pdf, (Accessed on July 17, 2020).

[79] FDA-funded NEST program names Data Science Institute AI use case as demonstration project, 2018, http://www.acr.org/Media-Center/ACR-News-Releases/2018/FDA-NEST-Program-Names-ACR-DSI-Use-Case-as-Demo-Project, (Accessed on July 17, 2020).

[80] A.P. Brady, Error and discrepancy in radiology: Inevitable or avoidable?, Insights Imaging 8(1) (2017) 171-182, https://doi.org/10.1007/s13244-016-0534-1.
A Multisite, Report-Based, Centralized Infrastructure for Feedback and Monitoring of Radiology AI/ML Development and Clinical Deployment

[81] M.I. Razzak, S. Naz, A. Zaib, Deep learning for medical image processing: Overview, challenges and the future, in: N. Dey, A. Ashour, S. Borra (Eds.) Classification in bioapps: Automation of decision making, Springer International Publishing, 2017, pp. 323-350, https://doi.org/10.1007/978-3-319-65981-7_12.

[82] E. Denham, Rfa0627721 – provision of patient data to DeepMind, 2017, http://ico.org.uk/media/action-weve-taken/undertakings/2014353/undertaking-cover-letter-revised-04072017-to-first-person.pdf, (Accessed on July 17, 2020).

[83] Information sharing agreement Royal Free Google UK, 2015, http://drive.google.com/file/d/0BwQ4esYYFC04NFVTRW12TTFFRF/view, (Accessed on July 17, 2020).

[84] J. Kirby, Crowds cure cancer: Data collected at the RSNA 2018 annual meeting - TCIA DOIs http://wiki.cancerimagingarchive.net/display/DOI/Crowds+Cure+Cancer%3A+Data+collected+at+the+RSNA+2018+annual+meeting, (Accessed on July 17, 2020).

[85] E. Ziegler, T. Urban, D. Brown, J. Petts, S.D. Pieper, R. Lewis, C. Hafey, G.J. Harris, Open health imaging foundation viewer: An extensible open-source framework for building web-based imaging applications to support cancer research, JCO Clin Cancer Inform 4 (2020) 336-345, https://doi.org/10.1200/CCI.19.00131.

[86] Y. Zheng, M. Wu, E. Cole, E.D. Pisano, Online annotation tool for digital mammography, Academic Radiology 11(5) (2004) 566-572, https://doi.org/10.1016/s1076-6332(03)00726-8.

[87] S. Ørting, A. Doyle, A. van Hilten, M. Hirth, O. Inel, C.R. Madan, P. Mavridis, H. Spiers, V. Cheplygina, A survey of crowdsourcing in medical image analysis, arXiv preprint arXiv: 1902.09159, 2019, (Accessed on July 17, 2020).

[88] N. Petrović, G. Moyà-Alcover, J. Varona, A. Jaume-i-Capó, Crowdsourcing human-based computation for medical image analysis: A systematic literature review, Health Informatics Journal (2020) 1460458220907435, https://doi.org/10.1177/1460458220907435.

[89] D.L. Weiss, C.P. Langlotz, Structured reporting: Patient care enhancement or productivity nightmare?, Radiology 249(3) (2008) 739-747, https://doi.org/10.1148/ radiol.2493080988.

[90] D. Pinto Dos Santos, S. Brodehl, B. Baessler, G. Arnhold, T. Dratsch, S.H. Chon, P. Mildenberger, F. Jungmann, Structured report data can be used to develop deep learning algorithms: A proof of concept in ankle radiographs, Insights Imaging 10(1) (2019) 93, https://doi.org/10.1186/s13244-019-0777-8.

[91] D. Clunie, Standardizing ai annotations the dicom way, 2018, http://www.dclunie.com/papers/CMIMI_2018_David_Clunie_StandardizingAIAnnotationsDICOM.pdf, (Accessed on July 17, 2020).

[92] S. Goldberg-Stein, V. Chernyak, Adding value in radiology reporting, JACR 16(9) (2019) 1292-1298, https://doi.org/10.1016/j.jacr.2019.05.042.

[93] G.B. McGinty, B. Allen, The ACR Data Science Institute and AI advisory group: Harnessing the power of artificial intelligence to improve patient care, JACR 15(3) (2018) 577-579, https://doi.org/10.1016/j.jacr.2017.12.024.

[94] B. McMahan, D. Ramage, Federated learning: Collaborative machine learning without centralized training data, 2017, http://ai.googleblog.com/2017/04/federated-learning-collaborative.html, (Accessed on July 17, 2020).
[95] P. Vepakomma, T. Swedish, R. Raskar, O. Gupta, A. Dubey, No peek: A survey of private distributed deep learning, arXiv preprint arXiv: 1812.03288, 2018, (Accessed on July 17, 2020).

[96] Intel works with University of Pennsylvania in using privacy-preserving AI to identify brain tumors. Intel newsroom, http://newsroom.intel.com/news/intel-works-university-pennsylvania-using-privacy-preserving-ai-identify-brain-tumors/, (Accessed on July 17, 2020).

[97] M.J. Sheller, G.A. Reina, B. Edwards, J. Martin, S. Bakas, Multi-institutional deep learning modeling without sharing patient data: A feasibility study on brain tumor segmentation, in: A. Crimi, S. Bakas, H. Kuijf, F. Keyvan, M. Reyes, T. van Walsum (Eds.), Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries. Brainles 2018, 2019, pp. 92-104, https://doi.org/10.1007/978-3-030-11723-8_9.

[98] T.M. Deist, F.J.W.M. Dankers, P. Ojha, M. Scott Marshall, T. Janssen, C. Faivre-Finn, C. Masciocchi, V. Valentini, J. Wang, J. Chen, Z. Zhang, E. Spezi, M. Button, J. Jan Nuyttens, R. Vernhout, J. van Soest, A. Jochems, R. Monshouwer, J. Bussink, G. Price, P. Lambin, A. Dekker, Distributed learning on 20 000+ lung cancer patients – the personal health train, Radiother Oncol 144 (2020) 189-200, https://doi.org/10.1016/j.radonc.2019.11.019.

[99] R.C. Mayo, L.Q. Chang Sen, J.W.T. Leung, Financing artificial intelligence in medical imaging: Show me the money, JACR 17(1) (2020) 175-177, https://doi.org/10.1016/j.jacr.2019.07.004.

[100] A. Alexander, A. Jiang, C. Ferreira, D. Zurkiya, An intelligent future for medical imaging: A market outlook on artificial intelligence for medical imaging, JACR 17(1) (2020) 165-170, https://doi.org/10.1016/j.jacr.2019.07.019.