Compiled Public Comments on

Request for Information: Critical resource gaps and opportunities to support radiological tool development and clinical data interpretation using artificial intelligence (AI)/machine learning (ML)

Guide Notice Number: NOT-OD-21-163

August 3, 2021 – November 1, 2021

Table of Contents

ID: 1698

Submit date: 9/18/2021

I am responding to this RFI: On behalf of myself

Name: Orhan K. Oz, MD, PhD

Name of Organization: UT Southwestern Medical Center

Type of Organization: Academic Institution

Role: Clinician

Domain of research most important to you or your organization (e.g. cognitive neuroscience, infectious epidemiology):

Bone health assessment

1. **Development of reference datasets, tools, and infrastructure to support radiological imaging analysis and interpretation using AI/ML (limit: 8000 characters)**

   There is a lack of normal bone density reference values from CT of the chest, abdomen or pelvis that could be used for opportunistically identifying low bone mass and quality. The reference data should be gender and ethnicity specific since bone mass varies according to these parameters. The existence of this reference dataset could help evaluate AI tools capabilities for determining bone mass and quality from CT scans.

2. **Existing Resources that could be leveraged to fill resource gaps (limit: 8000 characters)**

   I don't know for sure. Perhaps UCSF has some scans.

3. **Any general comments related to critical resource gaps and opportunities to support radiological tool development and clinical data interpretation using AI/ML (limit: 8000 characters)**
   Billions of dollars are spent each year related to bone health morbidities. Millions of CT scans are done each year in multiple health systems across the country. This is an opportunity to invest in a solution to a common problem.

ID: 1771

Submit date: 10/26/2021

I am responding to this RFI: On behalf of myself

Name: Daniel Barboriak

Name of Organization: Duke University School of Medicine

Type of Organization: Academic Institution

Role: Investigator/Researcher

Domain of research most important to you or your organization (e.g. cognitive neuroscience, infectious epidemiology):

Multi-center clinical trials in patients with brain tumors

1. **Development of reference datasets, tools, and infrastructure to support radiological imaging analysis and interpretation using AI/ML (limit: 8000 characters)**

   I am involved in the design and implementation of imaging for multicenter clinical trials for brain tumors. My responses here will focus on informatics needs for these trials, based on progress already made. My hope is that some of these comments will be applicable for those devising trials for non-CNS malignancies. We are very much in debt to those who supplied MRI cases and segmentation for BraTS challenges (https://pubmed.ncbi.nlm.nih.gov/25494501/). Thanks to these researchers, we now have many segmentation algorithms available. Recent publications of "adversary images" suggest that algorithms can be fooled by the alterations of the images in a manner not detectable by humans. This decreases confidence in whether these algorithms are sufficiently robust for use in clinical trials or practice. A pressing need to advance research here is to make imaging datasets to directly measure the variability of segmentation results widely available. Traditionally, this is accomplished using "coffee break" repeat imaging, and publishing more of these datasets would be helpful. One challenge is that there is uncertainty whether when a feature such as tumor-related enhancement is measured, that it is essentially the same object at both repeat time points. Research from our lab suggests that algorithms find a higher volume of tumor-related enhancement on later scans than earlier scans when these are obtained on the order of 48 hours apart. There may be a role in synthetic images to measure this uncertainty, if this approach can be validated. There may be a role for using sets of scans that trained observers assess as unchanged, even if further apart in time than traditional coffee break experiments. In this regard, I would note that neuroradiologists as a rule are not formally trained to draw free-hand outlines around tumor, so it is unclear how well that skill can be relied on. Neuroradiologists are, however, trained to detect subtle changes in images. Having imaging sets that are annotated using this training/skill as part of a BraTS challenge for use in training a change detection algorithm (e.g. using Siamese networks) would undoubtedly be valuable. A focus on collections of imaging data, particularly repeat datasets, unchanged datasets or related simulations using the standardized Brain Tumor Imaging Protocols for gliomas, metastases and CNS lymphomas (https://pubmed.ncbi.nlm.nih.gov/26250565/, https://pubmed.ncbi.nlm.nih.gov/32048719/, https://pubmed.ncbi.nlm.nih.gov/33560416/) would speed translation of algorithms into the clinical trials environment. A database of MR images that

expert raters would agree show no significant tumor-related enhancement would be helpful to address this important question: how reliable are AI/ML algorithms for calling "CR" -- complete response? Further development of lesion simulation approaches (https://doi.org/10.1117/12.2315704) should be encouraged, but more work is needed to validate these in comparison to actual patient data (e.g. coffee break experiments).

2. **Existing Resources that could be leveraged to fill resource gaps (limit: 8000 characters)**

3. **Any general comments related to critical resource gaps and opportunities to support radiological tool development and clinical data interpretation using AI/ML (limit: 8000 characters)**
   More progress is needed to answer some basic practical questions. For a particular disease (e.g. glioblastoma) which conventional MR imaging metric best reflects disease burden: cross sections, or volumes? Can we better triage treatments in early phase clinical trials by using changes in the rate of tumor volume growth rather than traditional progression to measure success? A recent paper discusses using this approach (https://pubmed.ncbi.nlm.nih.gov/34570454/), but this approach needs validation before wide adoption. In this context, expert agreement on standardization of methods to measure cross-sections using open source, widely available software would be helpful, perhaps building on prior approaches (https://pubmed.ncbi.nlm.nih.gov/31190077/). This software could then be evaluated using legacy datasets to evaluate whether software performed similarly to human raters in defining progression and response. Now that the imaging acquisition has been standardized using the Brain Tumor Imaging Protocol and we have multiple brain segmentation algorithms available as Docker containers (https://pubmed.ncbi.nlm.nih.gov/32410929/), there is a need for algorithms that assess tumor status (progression vs. stable disease vs. partial response vs. complete response) in an automated fashion. These algorithms should identify the areas of growth, and determine if there is no progression whether the current imaging should serve as a new baseline. The need to exclude non-tumoral enhancement evolving from areas of diffusion abnormality that appear after surgery must be addressed. Approaches have been published (https://pubmed.ncbi.nlm.nih.gov/30952559/) but wider availability of multiple open source algorithms in this space should be encouraged. Comparison of results of various algorithms on legacy datasets and correlation with patient outcomes would be helpful; if the software performs equally well as trained neuroradiologist raters, this would justify use of these algorithms in clinical trials and ultimately in practice. As a separate issue, there is a specific challenge we experience in evaluating intracranial metastatic melanoma, because many of the metastases are hemorrhagic, and demonstrate intrinsic hyperintensity on T1-weighted MR images. There would be value in determining what drives poor survival in these patients: number of metastases? Volume of enhancement? What is the significance of the appearance of non-enhancing hemorrhages suggesting underlying metastatic disease?

Email: daniel.barboriak@duke.edu

ID: 1775

Submit date: 11/1/2021

I am responding to this RFI: On behalf of an organization

Name: Tara Burke

Name of Organization: Association for Molecular Pathology

Type of Organization: Professional Organization/Association

Role: Organizational Official

Domain of research most important to you or your organization (e.g. cognitive neuroscience, infectious epidemiology):

Molecular Pathology.

1. **Development of reference datasets, tools, and infrastructure to support radiological imaging analysis and interpretation using AI/ML (limit: 8000 characters)**

   Please see the attached comments from the Association for Molecular Pathology (AMP). Thank you for the opportunity to comment.

2. **Existing Resources that could be leveraged to fill resource gaps (limit: 8000 characters)**

3. **Any general comments related to critical resource gaps and opportunities to support radiological tool development and clinical data interpretation using AI/ML (limit: 8000 characters)**

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/rfi2021_radiomics/uploads/NWJunxqmWQ.pdf

Description: Please see the attached comments from the Association for Molecular Pathology (AMP). Thank you for the opportunity to comment.

Email: tburke@amp.org

ID: 1776

Submit date: 11/1/2021

I am responding to this RFI: On behalf of an organization

Name: Katherine Pizer

Name of Organization: Medical Imaging and Data Resource Center (MIDRC)

Type of Organization: Other

Type of Organization-Other: Funded Center

Role: Research Participant/Patient/Advocate

Domain of research most important to you or your organization (e.g. cognitive neuroscience, infectious epidemiology):

AI/ML and ethically-sourced data commons for medical image analysis in disease detection, diagnosis, prognosis, treatment response monitoring, risk assessment, and patient management.

1. **Development of reference datasets, tools, and infrastructure to support radiological imaging analysis and interpretation using AI/ML (limit: 8000 characters)**

   As MIDRC, we have developed a large, public, diverse, clinically annotated, de-identified, and harmonized COVID-19 medical image repository, which also contains associated metadata, longitudinal data, demographics, and clinical records and variables. Though currently focused on COVID -19, our data model is extendable to other diseases and types of metadata, including genetics. The collection of such a large, ethically-sourced, trustworthy radiology image dataset that adheres to the FAIR principles is a monumental task for which we have been meticulously working through every step - from data transfer agreements to patient data de-identification to harmonization of data ingested from different clinical institutions to creating a user-interface for researchers from across the globe, allowing them to build complex cohorts for their research studies. This current effort, led by the RSNA, ACR, and AAPM, is being conducted with over 100 investigators from 24 institutions including academic medical centers, hospitals, private practices, and FDA. As a multi-institutional initiative comprised of a wide variety of imaging stakeholders and domain experts, research is also an important focus of our work and our investigators bring a wealth of expertise to the table in radiology AI/ML applications, radiomics and radiogenomics research, performance assessment, and equity, diversity, and inclusion. Within our current focus on COVID-19, multiple in-house radiology AI/ML studies are being undertaken, scientific Challenges are being planned, publicly-available performance evaluation tools are being developed, and creation of a sequestered commons will all benefit the scientific community at-large. Two of the gaps impeding high-quality reproducible research are 1) the lack of diverse, trustworthy datasets for AL/ML researchers and 2) the lack of mechanisms for researchers to have their algorithms evaluated and validated, either for research purposes or regulatory approval. Reducing the barriers to access to high-quality, de-identified, curated, annotated data levels the playing field, not only for training algorithms, but also for benchmark testing of different algorithms' performance to enable translation to the clinic to benefit the public good. To date, MIDRC has released over 15,000 imaging

studies to the public, with another 48,000 ingested and currently undergoing quality control prior to release. A single data commons to which multiple clinical sites can contribute aims to eliminate the existence of Frankenstein datasets, which consist of cases from multiple repositories with a potential for duplicative cases that cause leakage between training and testing datasets. Our representative cloud-based repository includes both an open commons and a sequestered commons; both of which can serve as national resources. The sequestered commons includes cases that will never be shared but will be used to independently test AI/ML algorithms to expedite translation through the regulatory process. The third gap in translational radiology AI/ML research is the high threshold for clinical sites to contribute data to a data commons. De-identification of radiology exams and associated metadata is a large responsibility and a clinical site interested in data contribution to a repository faces many hurdles, from obtaining IRB approval to having the required hardware/software to specialized personnel needs. This is especially true for smaller clinical sites in under-served areas, leading to a potential population bias in the data collected and further bias in the AI/ML algorithms developed, unless this bias is addressed, accounted for and mitigated, to the extent possible. There is also a lack of incentive for clinical sites to contribute data other than to serve the common good. Funding should be made available not only for the creators of image data repositories themselves, but for contributing clinical sites. Another gap impeding clinical and translational research is the difficulty in obtaining clinical data associated with radiology exams (those are often stored in different systems within a clinical site) and linking those to the appropriate exams and exam time points. De-identification needs to be performed in a way that preserves each patient's longitudinal data and timeline, an especially significant issue for data obtained from multiple clinical sites that may pertain to one patient. We are also working to resolve all related and limiting issues that arise when information is linked and shared between multiple data repositories, for example between ours and the National COVID Cohort Collaborative (N3C) as well as with NHLBI BioData Catalyst.

2. **Existing Resources that could be leveraged to fill resource gaps (limit: 8000 characters)**

   As a multi-institutional initiative comprised of academia, imaging societies, government agencies and industry partners, we are currently establishing our two-commons system of open and sequestered datasets, and developing new diagnostics and AI/ML algorithms. We are further able to leverage each of our collaborators' existing expertise within the medical imaging community, their processes and their infrastructure; infrastructure that already meets accepted standards of quality, security, access and sustainability. We serve as a linked data commons that harmonizes access to data and data management activities across all participating organizations at three critical stages: (1) intake, including curation, de-identification, harmonization, and quality assessment, (2) annotation and labeling of images and other data using semi-automated approaches, and (3) distributed access and query methods. To date, we have more than 15,000 COVID-related imaging exams available for researchers to download (data.midrc.org) and have ingested another 48,000 cases in our pipeline. As it pertains to this RFI, we can serve as a national resource for multiple medical imaging needs.

3. **Any general comments related to critical resource gaps and opportunities to support radiological tool development and clinical data interpretation using AI/ML (limit: 8000 characters)**
   Inappropriate or incomplete statistical analysis, as well as the lack of reproducibility of research studies, are known hurdles in validation of AI/ML algorithms and their translation to clinical practice. Another often-overlooked aspect is the presence of bias in AI/ML algorithms, either because data

collected and used was not representative of the population, the human 'truthing' process was biased, or because of errors in the research methods themselves. As such, the investigation, prevention, and mitigation of bias in our data commons and subsequently developed AI/ML algorithms is one of our core missions. Investigators of MIDRC (Medical Imaging and Data Resource Center) are specifically interested in a) bias assessment and mitigation in medical imaging (such as chest radiographs in COVID-19) combined with clinical (electronic health record), demographic, and image acquisition modality data (e.g., scanner model), b) establishing a fairness metric and bias score to mitigate bias in AI development, c) developing unbiased algorithms and investigating such de-biased AI algorithms at scale to understand root causes of health disparities, and d) investigating the impact of biased algorithms on clinical practice.

Email: kpizer@bsd.uchicago.edu

ID: 1778

Submit date: 11/1/2021

I am responding to this RFI: On behalf of an organization

Name: Curtis P. Langlotz, MD, PhD

Name of Organization: Stanford University Center for Artificial Intelligence in Medicine and Imaging

Type of Organization: Academic Institution

Role: Investigator/Researcher

Domain of research most important to you or your organization (e.g. cognitive neuroscience, infectious epidemiology):

Development of machine learning methods to optimize how imaging and other clinical data are used to improve human health

1. **Development of reference datasets, tools, and infrastructure to support radiological imaging analysis and interpretation using AI/ML (limit: 8000 characters)**

   We have proposed treating health data as a "public good" and called for responsible secondary use of health data [1]. The growing momentum behind these concepts have spawned efforts to make anonymized medical data available to the open science and education communities [2]. However, several gaps remain: Gap #1: Dataset Readiness Models learned from healthcare data are often not useful, reliable, and fair [3], leading to significant research on algorithmic fairness [4]. Most available datasets suffer from a lack of diversity [5], and a paucity of high-quality labels necessary for machine learning, including the diagnosis, demographics, and other critical elements of clinical context. Consequently, only a small fraction of open medical datasets are "AI-ready." For data to be useful for machine learning, labeling and annotation of medical images is required, either by extracting labels from the electronic health record or by acquiring labels from expensive human medical experts. Even experts with years of clinical training are affected by significant inter-observer variability. As a result of these shortcomings, only a small fraction of imaging data are annotated in an "AI-ready" format. For example, the lack of "AI readiness" has also affected the development of AI applications to address COVID-19. Many open COVID medical datasets lack critical information, have errors in curation and annotations, and incorporate clinical/socioeconomic biases, leading to machine learning solutions that were biased or not useful in practice [6, 7]. Gap #2: Organizational Readiness for Data Sharing The NIH Policy for Data Management and Sharing is on a collision course with the current capacity of health research organizations. The spirit of the NIH policy is to create a health data ecosystem where all motivated data scientists have access to the data, computation, and translation resources they need to develop, test, and implement their innovations to benefit society. Unfortunately, organizations that are the sources of health care datasets face numerous obstacles to the wide release of data: 1. Risk aversion. Health care organizations face significant financial penalties and possible damage to their reputation if a data breach occurs. 2. Organizational structure. Smaller healthcare organizations and organizations that care for underserved populations may have no IRB and no privacy office. Larger organizations often have IRBs, but can comprise multiple organizations, each of which may assert authority over privacy issues. 3. Legal costs.

Expensive legal deliberations are often required to develop patient consent forms and data use agreement that provide the flexibility that researchers need, while protecting the rights of the patients and releasing organizations. 4. Technical factors. To assure patient privacy and manage risk, released data must be de-identified using a complex automated process, followed by a manual process to confirm that protected health information (PHI) has been removed. Data hosting systems are costly to produce and customize. Gap #3: Lack of Platforms for Hosting and Dissemination of Imaging Data Most current imaging data repositories support specific use cases, such as cancer (e.g., TCIA, Cancer Imaging Data Commons) or COVID-19 (e.g., MIDRC). Some imaging repositories lack key capabilities necessary to ingest, process, store, and disseminate AI-ready data, such as: (1) harmonization of study descriptions and other non-structured meta-data across sites, (2) acquisition and storage of annotations, (3) sequestration methods that enable the benchmarking of machine learning tools for regulatory or other purposes, (4) honest brokers to enable linkage of imaging data to repositories containing other forms of data, such as BioData Catalyst and N3C, and (5) tools for researchers to select useful cohorts based on imaging meta-data. Gap #4: Tools to Reduce Mismatch between Available Datasets and Clinically Important Problems There are many important diseases for which little or no AI-ready imaging data is widely available: 1. Researchers need data that reflect real-world complex mix of patient conditions and comorbidities. 2. There are extremely few raw datasets available to improve image reconstruction algorithms. 3. Clinical trial data and other real-world evidence (clinical care data) is often inaccessible and unlabeled. 4. Few datasets contain imaging studies showing the progression or regression of disease over time. Clinical care represents a dynamic scenario in which health professionals and patients have clinical encounters across time. Often clinicians use information from multiple time points to improve diagnostic and prognostic accuracy. This is particularly true for imaging tasks, where comparisons are often performed to detect changes and to understand the disease state at any point in time. Gap #5: Lack of Evaluation Frameworks to Assess Usefulness, Reliability and Fairness Using a machine learning model to improve care is an interplay between: (1) the model and the risk estimate it provides, (2) the workflow and policy for allocating an intervention at a certain risk level, and (3) the benefit-harm trade-off of the intervention. There is a lack of research on best practices for assessing fairness, reliability, and usefulness of model-guided interventions. For example, a risk stratification model may fail because: (1) The outcomes a model predicts are refractory to interventions. (2) The prevalence of a condition may be so low that, even at high precision, the number of false positives overwhelms the workflow. (3) A model can suffer from its own success when its use drives actions that improve outcomes so much that the performance of the model appears to degrade. (4) Disparities can be exacerbated for disadvantaged populations when inequities in patterns of care are institutionalized in a model, or when populations are underrepresented in the datasets used to train the models, or when biased or mis-specified proxy outcomes are used during model development, leading to differences in the performance of models across subgroups. The need for measures of usefulness, reliability and fairness has led to multiple recommendations (see Model Cards, MI-CLAIM, MINIMAR, TRIPOD, TRIPOD-AI, STARD, STARD-AI, QUADAS-AI, SPIRIT-AI and CONSORT-AI) but adherence to recommendations remains limited and concrete best practices are lacking [8]. References 1. Larson DB, Magnus DC, Lungren MP, Shah NH, Langlotz CP. Ethics of Using and Sharing Clinical Imaging Data for Artificial Intelligence: A Proposed Framework. Radiology. 2020;295: 675–682. 2. Irvin J, Rajpurkar P, Ko M, Yu Y, Ciurea-Ilcus S, Chute C, et al. CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison. arXiv [cs.CV]. 2019.

http://arxiv.org/abs/1901.07031 3. Striving for Diversity in Research Studies. N Engl J Med. 2021. doi:10.1056/NEJMe2114651 4. Chen IY, Pierson E, Rose S, Joshi S, Ferryman K, Ghassemi M. Ethical Machine Learning in Healthcare. Annu Rev Biomed Data Sci. 2021;4: 123–144. 5. Kaushal A, Altman R, Langlotz C. Health Care AI Systems Are Biased. Scientific American. 17 Nov 2020. Available: https://www.scientificamerican.com/article/health-care-ai-systems-are-biased/. Accessed 30 Sep 2021. 6. Wynants L, Van Calster B, Bonten MMJ, Collins GS, Debray TPA, De Vos M, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ. 2020;369: m1328. 7. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366: 447–453. 8. Petersen C, Smith J, Freimuth RR, Goodman KW, Jackson GP, Kannry J, et al. Recommendations for the safe, effective use of adaptive CDS in the US healthcare system: an AMIA position paper. J Am Med Inform Assoc. 2021;28: 677–684.

2. Existing Resources that could be leveraged to fill resource gaps (limit: 8000 characters)

Opportunity #1: Capitalize on the Expertise of Data Release Centers of Excellence Organizations need a roadmap and a set of best practices to navigate the obstacles to data release, including examples of dataset releases that have successfully managed privacy and consent issues, data use agreements that are acceptable both to the releasing organization and the data user, and software that can overcome the technological barriers to data release. There is also a need to develop clear standards for optimal characteristics of useful datasets, including clinical utility, size and scale, fairness, annotation and labeling methods, data diversity, the need for a held-out test set, and other factors. When paired with the appropriate resources, the NIH vision of a race to the top can be achieved. Research centers such as our own, the Center for Artificial Intelligence in Medicine and Imaging, have become centers of excellence in the wide release of imaging data. We have released 10 large clinical datasets. We would welcome the opportunity to serve as a resource, sharing our knowledge and experience establishing organizational and data readiness with less experienced sites. Other data release sites, such as Nightingale Open Science and MIMIC could also serve as resources. Opportunity #2: Reduce the Image Labeling Burden Significant progress has been made toward methods that can dramatically reduce labeling burden. For example, contrastive learning methods between images and text descriptions, which are available for many image types, can reduce labeling needs by 1-2 orders of magnitude [9, 10]. Active learning methods can select the optimal study to label next, further reducing labeling burden [11, 12]. Weak labeling, semi-supervised learning, and unsupervised learning methods, which have been effective in reducing labeling in other domains, are increasingly being formulated for imaging tasks [13]. Generative adversarial networks, including physics-based data generation methods, produce data augmentation methods that reduce the need for manually labeled data [14]. Opportunity #3: Repurpose and Expand the Infrastructure of Current Data Platforms MIDRC is a large, diverse, public, annotated, de-identified, and harmonized medical image repository, focused on COVID-19. The MIDRC repository also contains associated metadata, longitudinal data, demographics, and clinical records and variables. This effort, led by the RSNA, ACR, and AAPM, requires the development of data transfer agreements, patient data de-identification methods, quality assurance and harmonization of data ingested from a wide variety of healthcare organizations, and user-interfaces enabling researchers from across the globe to build complex cohorts for their research studies. To date, MIDRC has made more than 15,000 curated COVID-related imaging exams available for researchers to download

(data.midrc.org) and has ingested an additional 48,000 cases. Though currently focused on COVID - 19, with the appropriate support, the data model is extendable to other diseases, imaging modalities, and types of metadata. Likewise, secure linkages to repositories containing other forms of data require resources for support. MIDRC could be sustained and repurposed to become an ethically-sourced, trustworthy radiology image dataset across diseases and imaging modalities that adheres to the FAIR principles.

3. **Any general comments related to critical resource gaps and opportunities to support radiological tool development and clinical data interpretation using AI/ML (limit: 8000 characters)**
Overwhelming evidence now shows that machine learning models do not generalize well to new settings [15]. This not only argues for the production of diverse multi-institutional datasets as described above, but also indicates the need for analytic methods that can monitor the performance of algorithms over time. Benchmark datasets and post-market analytics could be developed to support new regulatory models, and could produce revenue to help sustain imaging repositories over time. Several new research methods are designed to preserve patient privacy while making large datasets widely available to train machine learning models. These methods, including federated learning, synthetic data, and differential privacy, often require the close coordination between sites or the tailoring of mathematical methods to specific use cases. These methods could be explored scientifically, but are not yet ready to replace the function of centralized repositories.
References 9. Zhang Y, Jiang H, Miura Y, Manning CD, and Langlotz CP. "Contrastive Learning of Medical Visual Representations from Paired Images and Text." Proceedings of the International Conference on Learning Representations (ICLR 2021, acceptance rate 29%) https://arxiv.org/abs/2010.00747 10. Chen T, Kornblith S, Norouzi M, Hinton G. A Simple Framework for Contrastive Learning of Visual Representations. http://arxiv.org/abs/2002.05709 11. Wei Q, Chen Y, Salimi M, Denny JC, Mei Q, Lasko TA, et al. Cost-aware active learning for named entity recognition in clinical text. J Am Med Inform Assoc. 2019. doi:10.1093/jamia/ocz102 12. Wei Q, Chen Y, Salimi M, Denny JC, Mei Q, Lasko TA, et al. Cost-aware active learning for named entity recognition in clinical text. J Am Med Inform Assoc. 2019. doi:10.1093/jamia/ocz102 13. Hooper S, Wornow M, Seah YH, Kellman P, Xue H, Sala F, et al. Cut out the annotator, keep the cutout: better segmentation with weak supervision. ICLR, 2020. Available: https://openreview.net/pdf?id=bjkX6Kzb5H 14. Sorin V, Barash Y, Konen E, Klang E. Creating Artificial Images for Radiology Applications Using Generative Adversarial Networks (GANs) – A Systematic Review. Acad Radiol. 2020. doi:10.1016/j.acra.2019.12.024 15. Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. PLoS Med. 2018;15: e1002683.

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/rfi2021_radiomics/uploads/SduybwEumF.pdf

Description: MS Word version of the comments above

Email: langlotz@stanford.edu

ID: 1781

Submit date: 11/1/2021

I am responding to this RFI: On behalf of an organization

Name: Michael Boss

Name of Organization: American College of Radiology

Type of Organization: Non-Profit Research Organization

Role: Investigator/Researcher

Domain of research most important to you or your organization (e.g. cognitive neuroscience, infectious epidemiology):

Imaging of cancer, integrated diagnostics and advanced imaging analytics

1. **Development of reference datasets, tools, and infrastructure to support radiological imaging analysis and interpretation using AI/ML (limit: 8000 characters)**

   The American College of Radiology (ACR) works closely with ECOG-ACRIN and members at associate institutions to advance cancer diagnosis and treatment to better serve patients. We see tremendous potential to advance patient care by further development of integrated diagnostics. This effort can combine rich imaging, pathology, genetic, and clinical datasets to create new diagnostic tools that leverage multiple datastreams. These same datasets can also serve a foundational starting point for advanced metrics and analytics: radiomic, pathomic, and genomic feature sets can inform predictive and classification models that can better identify disease or disease subtype, predict disease progression, and potentially identify optimal treatment regimens. We see opportunity to provide access to reference datasets and harmonized tools in both centralized and distributed platforms. These data and tools can aid the medical imaging community to develop robust analytic techniques that are cross-comparable using standardized metrics. Researchers will benefit from lower barriers to entry, greater access to data, and provision of tools that allow for more advanced analyses, including the use of machine and deep learning algorithms for greater predictive and classification accuracy. Data must be made more accessible, which will require thorough and automated anonymization during the data preparation phase. These efforts should include imaging, clinical, and other relevant datatypes to maximize the potential value of retrospective and prospective studies, while ensuring correlation across multiple datatypes and timepoints within a particular study. Furthermore, there is a need for appropriate infrastructure to enable on-study analyses, utilizing multiple datastreams from a study to guide decision-making. For example, the use of real-time radiogenomic analyses could improve screening performance, inform treatment approach, and predict outcomes, allowing for more informative study design. Imaging data can achieve more of its potential through improved curation, thereby enabling advanced analytical workflows to facilitate the development, training, and validation of machine and deep learning algorithms. Imaging studies should not only be added to public repositories, but parsed and interrogable by anatomical coverage, series names and descriptors, and disease and treatment regimen. Series names and descriptors should adhere to standardized lexicons to facilitate their use in search queries. When possible, any existing segmentations and annotations should be included as part of imaging

datasets. When absent, tools to segment and annotate, especially in an automated fashion, should be employed to add value and utility to images.

2. **Existing Resources that could be leveraged to fill resource gaps (limit: 8000 characters)**

   While there are efforts to create public repositories of data and tools, to date these have been heavily focused on retrospective studies. Efforts such as The Cancer Imaging Archive (TCIA) or Imaging Data Commons (IDC) provide access to such studies, and even include image viewers. However, these platforms do not provide an end-to-end pipeline of dataset identification, selection, viewing, segmentation, and analysis. The ACR's Data Access and Research Toolkit (DART) provides a centralized platform for data and tool hosting, while affording access to both retrospective and prospective studies. DART can host and provide access to imaging, genomic, pathology, clinical, and other data. Users request access to study data and then query these data to generate a curated dataset filtered by elements such as demographics, study arm, etc. DART supports the incorporation of analytical tools, e.g., through containerization. DART reduces the barrier to data access and tool use, while simultaneously providing a distribution channel for tool developers. The ability to facilitate this analysis provides researchers greater opportunities to focus their attention on advanced analytics and tool development without having to build local pipelines at their institutions, while also benefitting from access to multicenter data. DART has incorporated several automated analysis pathways for particular use cases (such as phantom analysis), as well as more generalized tools for segmentation and generation of radiomic feature sets. Recently, ACR staff and researchers from the University of Pennsylvania have worked to integrate the Cancer Imaging Phenomics Toolkit (CaPTk) suite of tools into DART, enabling advanced image analysis of DART-hosted data within the DART platform. The standalone instance of CaPTk can view a wide variety of imaging modalities, provide semi-automated segmentation, and generate radiomic feature sets. By early 2022 the DART-hosted instance of CaPTk will allow users to output these artifacts and store them for later use and sharing. Other examples of tools hosted or to-be-hosted in DART include the QIN-supported PyRadiomics toolbox, Orange, and Jupyter Notebooks. ACR Connect, installed on premises, brings diverse medical imaging analysis tools to site users through a series of applications operating on data that remains at their facility. Without the need for data to leave site firewalls, ACR-Connect applications offer utilities such as participation in cross-site research initiatives and creation of curated local datasets. AI-LAB, one ACR-Connect application, provides non-technical users the ability to create or evaluate artificial intelligence (AI) algorithms through an intuitive graphical interface. The use of AI-LAB can expedite the development of site-specific AI algorithms fine-tuned on local data through transfer learning for tasks such as disease classification or predictive modeling. Beginning in early 2022, AI-LAB users will be able to participate in federated learning, a method of collaborative AI algorithm training which does not require data sharing. AI-LAB will allow sharing of AI algorithms with these participating sites to analyze their own local data. Algorithms created through federated learning will be more robust and translatable across sites by virtue of exposure to training data from diverse institutions and patient populations. Focused as much on educating the radiology community as on developing robust AI models, AI-LAB offers videos explaining important AI concepts, presents model results in interactive figures, and provides thorough help content throughout the application.

3. **Any general comments related to critical resource gaps and opportunities to support radiological tool development and clinical data interpretation using AI/ML (limit: 8000 characters)**

Significant challenges impede progress towards the goal of integrated diagnostics, including the need for improved access to multicenter data and greater adoption of advanced analytic tools and metrics. Different datastreams often reside at different institutions, despite being from the same study. Accessing data from studies (such as those within the NCTN) can be challenging, as data access requests often require extensive paperwork and time, as well as various levels of review prior to approval. The necessary de-identification of patient data adds another hurdle that limits data sharing between institutions. Implementation of analytic tools requires significant resources, and often results in duplicative effort across sites. Harmonized tools are rarely available at the point of care. Advanced artificial intelligence (AI) algorithms are often trained using local, single-institution data, which can result in fragile findings that do not hold up when exposed to novel data from multiple sites. There are opportunities to build and expand both DART and ACR-Connect infrastructure to make more data available to researchers. Tools hosted in DART, such as CaPTk, need to be able to generate outputs that can be passed on to other tools within DART. This ability will create a self-contained analysis pipeline construction kit and allow for powerful data processing to greatly accelerate the development of artificial intelligence algorithms for diagnosis, treatment decisions, and prediction of outcomes. Although a standardized approach to containerization and integration of tools is necessary to expedite tool integration and to expand the variety of tools available to users, containerization of CaPTk effectively demonstrates proof-of-concept in the capability of DART to integrate externally developed tools. Additionally, expanding ACR-Connect support to datatypes beyond medical imaging would provide site users the ability to curate corresponding multitype datasets. This would further enhance the potential of applications such as AI-LAB to generate useful AI algorithms. Increasing access to radiological data will, in turn, enable faster analytic gains from the research community. The ACR has gained much experience in scaling up and automating anonymization processes in handling COVID-19 data as part of the Medical Imaging Data Resource Center (MIDRC). Many of the lessons learned have application with the wider body of imaging and clinical data and can serve to accelerate access to these data by the research community, while ensuring patient privacy and rigorous anonymization. Still other opportunities exist to develop infrastructure that will increase access to rich multitype data. For example, anonymized data from NCTN studies should be made available as public datasets once primary aims have been published: preparation for this availability should happen in parallel with analysis and preparation for publication so that external researchers can maximize the value of a study while it is most relevant, to better inform future trials and clinical care. Hosted safe harbor data should be accessible to interested users through automated processes that reduce the administrative overhead inhibiting rapid access to data under today's procedures for data access requests. The ability to perform harmonized analyses, either centrally (with DART) or in a distributed fashion (with AI-LAB), permits crowd-sourced approaches to data analysis, where appropriate. A critical area of need for machine/deep-learning segmentation algorithms is in generating accurate ground-truth segmentations. Creating challenges hosted centrally or distributed through a federated model could allow for crowd-sourced segmentations, which would then serve as ground-truth for AI algorithms. This approach can be applied to a wide variety of pressing AI problems, e.g., classification of breast density. An example of such data is the EAY131 "MATCH" study, where rich imaging, clinical, and genomic data exist. The ECOG-ACRIN Radiomics Working Group has requested and received access to this data; however, the group's members are distributed at many different institutions. A harmonized approach to analysis (e.g., segmentation and generation of radiomic

features) would greatly aid the research community while simultaneously adding tremendous value to the MATCH datasets for current and future analyses. Combining harmonized analysis and data preparation (e.g., implementation of ETL processes using OMOP Common Data Model tools) will yield even greater utility for studies with multiple datatypes and facilitate deeper analysis by diverse research teams. Harmonization of analysis within a study, integrated datasets, and improved access to data and tools are key needs of today's imaging landscape. Addressing these, in combination with expert input to add value to existing data in the form of ground-truth segmentations, annotations, classifications, and outcomes, will greatly accelerate the pace of automation from AI, the scope of its insights, and its utility and accuracy. It is paramount that these needs are addressed not in just one disease site or imaging modality, but in a comprehensive fashion that addresses the broad needs of the radiology and wider medical research communities.

Email: mboss@acr.org

ID: 1782

Submit date: 11/1/2021

I am responding to this RFI: On behalf of an organization

Name: Jesse Tetreault

Name of Organization: Nvidia

Type of Organization: Industry (Biotech/Device/Pharmaceutical Company)

Role: Investigator/Researcher

Domain of research most important to you or your organization (e.g. cognitive neuroscience, infectious epidemiology):

Accelerated Computing; GPU Computing; Medical Image Analysis

1. **Development of reference datasets, tools, and infrastructure to support radiological imaging analysis and interpretation using AI/ML (limit: 8000 characters)**

   "Scaling Computation as a Catalyst for Scientific Progress" Supporting and nurturing the adoption of Artificial Intelligence (AI) in both clinical and non-clinical applications of healthcare is imperative for advancing science in the modern era. Providing investigators and scientists sufficient computational power to formulate, test, and refine these new AI-driven hypotheses is the most important variable of progress that we are able to control, and is a direct catalyst for accelerating progress in every other dimension of AI that we care about. Progress in the field of AI - and in particular, Deep Learning - is driven by three fundamental factors: data, models/algorithms, and computation. Data - data quality, data quantity, dataset size, dataset variability, dataset bias - can directly constrain the possibilities of what most AI models can learn. If the goal is to develop a deep learning model in the context of radiology, e.g. detecting lung nodules in a chest X-ray, the dataset that the model was trained on should represent the setting in which it will be deployed for inference. It should also contain enough variability such that it can be generalized enough to be useful to real-world patients. The lack of a dataset that is of large enough magnitude or variability to train a useful model can often be an inhibitor to AI progress in radiology and related fields. Advances in DL model architecture and training techniques also directly drive progress in AI, sometimes quite rapidly. Several innovations in model architecture since the inception of deep learning circa 2012 - complex convolutional layers, residual networks, and more recently, the Transformer model - have unlocked completely new ways of framing problems and generating hypotheses. Even more remarkable and encouraging, several of these innovations often apply to more than one application of AI. The Transformer model that was originally developed for natural language processing is now regularly being adapted and trained to create state-of-the-art computer vision models for medical image analysis. The third factor that drives progress in AI is computation. NVIDIA GPUs have undoubtedly been the driving force behind almost every single significant advance in deep learning in the last decade. Although initially there were plenty of relatively large advances in AI due to the huge performance advantage of deep models over traditional approaches, training these models was extremely expensive. Deep learning software frameworks such as Caffe and Tensorflow enabled researchers to construct and train models more easily than coding them by hand, but lacked the

maturity to scale to the largest supercomputers of the time. Most of the time even the most sophisticated models were constrained to a single GPU, or at best a single node with two or four GPUs. NVIDIA has innovated to create new hardware, software, and system-level approaches to remove various computational bottlenecks created by AI workflows over the last decade. In 2016, when the popular ResNet model was first introduced, it had about 11 million learnable parameters and could be trained in a reasonable amount of time on a single system with eight or 16 GPUs. For perspective, to train the recent MT-NLG model for natural language processing created by NVIDIA and Microsoft, it can take over 420 DGX server nodes with eight GPUs each to reach an analogous accuracy in a reasonable amount of time. The MT-NLG model has over 500 billion learnable parameters and takes over 3,000 times more computing power to train. This is in only a five year time span, and is in addition to NVIDIA relentlessly optimizing the software and hardware stack to get incredible performance gains year after year. The fundamental reason for this is the ever-increasing capacity and complexity of deep learning models as we continue to push the frontier of artificial intelligence. We can expect this trend to continue. But, increasing computational resources for AI also has two other benefits: it accelerates model development and algorithm research by allowing researchers to experiment faster with new ideas, and it can also even help address data quality, quantity, or robustness issues. Generative models, particularly Generative Adversarial Networks, are able to learn how to generate new synthetic images that are so closely related to real images, they can supplement the original dataset. This allows us to approximate what corner cases may look like that we don't have data points for, e.g. such as a tumor in an odd or uncommon location, and help give the model an idea of what that may look like if it encounters such a data point in the real world. Investing in and building IT infrastructure that supports deep learning research is the best way to accelerate all facets of AI in radiology. In addition to increasing the raw number of experiments that can be done per day, sufficient and powerful computation also enables more creative model development, and can help address the common data shortage issues encountered in radiology with generative models trained on these same platforms. Providing an AI research team with the proper GPU infrastructure and software resources is analogous to handing them the keys to a seemingly limitless computational microscope.

2. **Existing Resources that could be leveraged to fill resource gaps (limit: 8000 characters)**

3. **Any general comments related to critical resource gaps and opportunities to support radiological tool development and clinical data interpretation using AI/ML (limit: 8000 characters)**
   Historically, the variety of computing needs across a department required purchases of specialized hardware and software for each intended use. The GPU is an accelerator that supports a wide range of uses within the data center, from High-Performance Computing, scientific visualization, deep learning training and inference, data science, imaging systems, and more. Generational advancements in GPU hardware, together with the integration of enterprise management software from segment leader VMware, delivers the ability to address those once disparate compute needs from a homogeneous computing platform. Enterprise virtualization enables administrators to provision data center hardware to end users based on their computational needs. Scientists developing Deep Learning algorithms where radiology images are an input for predicting potential patient issues need a robust accelerated computing platform with multiple GPUs. Other users require computing provisioned for smaller individual computing needs such as desktop analysis of DICOM images with 3D Slicer. Purchasing equivalently configured systems capable of simultaneously

executing multiple clinical tasks reduces data center complexity and increases computational availability. Utilization of the hardware can be further maximized through orchestration techniques using software that enables virtual desktops during the day and changes to enable HPC workloads during off-peak hours. These mixed workload uses are becoming increasingly common where on-premise data center facilities are nearing capacity since it reduces the amount of hardware required to meet the department's computing needs. A feature recently introduced by NVIDIA on data center compute GPUs supports a new method of GPU sharing called Multi-Instance GPU (MIG) This method of sharing GPU computing resources among multiple users simultaneously is similar to the existing process available with virtual machines provisioned using VMware, but instead of a temporal approach to sharing the GPUs (time slicing), a spatial partitioning option is available. The MIG approach supports users having a dedicated division of GPU computing resources rather than division of time (in milliseconds) using the GPU resources. For smaller workloads, this provides a more deterministic quality of service since MIG provides a set number of GPU cores for only one user. In addition, traditional Virtual Desktop Infrastructure (VDI) technology provides basic desktop needs for all, including high-end workstations used by research scientists, PACS imaging stations used by clinicians, and everyone who accesses EMR and EHR. To deliver a physical workstation experience from the data center, NVIDIA's Virtual GPU software is integrated with leading enterprise software from VMware, Citrix, and RedHat, providing clinicians with a GPU-accelerated workstation managed and delivered from the data center The variety of computing requirements in hospital settings presents challenges to IT administrators who must balance increasingly higher resolution data such as 3D images, expectations that staff can access the data securely from multiple locations on any approved device, and provide tools that improve the quality of care by enabling real-time collaboration among doctors and specialists. With enhanced productivity and mobility, doctors, nurses and staff can now access medical records and patient data on-the-go, on any device with a native-like experience. Radiologists and medical imaging specialists can also have remote access to large images for increased productivity. This rapid access to information for healthcare professionals enables faster decision-making and improved diagnostic accuracy. Furthermore, accessibility and ease of patient documentation results in improved clinical workflow. NVIDIA virtual GPU (vGPU) enables more predictable infrastructure costs by utilizing a common set of hardware capable of supporting the variety of computing requirements across the organization. IT can virtualize EMR and PACS applications, delivering them cost effectively to all users through the use of virtual machines. The staff become untethered to thick clients, supporting a variety of end-point devices like tablets, laptops, zero clients, and Chromebooks, without compromising the user experience. This simplification of infrastructure and its management reduces IT operations and maintenance costs while providing visibility across the entire virtualized infrastructure. Access to critical clinical applications while adhering to strict government regulations (HIPAA and HITECH) is maintained since records data is not transported to the end-point device, but rather it stays within the data center. These IT techniques benefit three key healthcare user groups: Radiologists and medical imaging specialists viewing and editing very large and complex medical images (PACS). Once tied to physical workstations, NVIDIA virtual GPU enables them to remotely access and edit very large and complex medical images using PACS so they can do supplemental diagnostic work at home or on the go, as opposed to being at radiology stations. Another use case is for remote collaboration - a specialist in one location can obtain the opinion of a remote specialist for their advice and opinion through voice, video, and records sharing. Radiologists, Specialists, and Clinician "Super

Users" who need to remotely view and edit medical images. These users remotely view and edit medical images so that they can, for example, show 3D CT scan results to patients on their iPads as they make their rounds. Doctors, Clinicials, Nurses, and staff who access virtualized EMR, billing, and office productivity applications. This group routinely uses virtualized EMR applications like EPIC to access medical records remotely and manage clinical workflows. They also use applications to better stay in touch and communicate with in-hospital patients by providing them with personalized materials and schedules based on access to test results and holistic patient information.

Email: jtetreault@nvidia.com