Clinician-driven AI in Ophthalmology: Resources Enabling Democratization

Authors: Edward Korot*[1,2], Mariana B. Gonçalves[2,3,4], Saad M Khan[5], Robbert Struyven[2,6], Siegfried K. Wagner[2], Pearse A Keane [2]

Affiliations:

1. Stanford University Byers Eye Institute, Palo Alto, USA
2. Moorfields Eye Hospital, London, United Kingdom
3. Federal University of São Paulo (UNIFESP), Sao Paulo, Brazil
4. Vision Institute (IPEPO), Sao Paulo, Brazil
5. Royal Berkshire Hospital, Reading, United Kingdom
6. University College London, London, United Kingdom

*Corresponding Author:
Edward Korot
2450 Watson Ct Palo Alto Ca 94303
(p)650-723-6995
(f)650-320-9443
ekorot@gmail.com

Abstract

Purpose of review

This article aims to discuss the current state of resources enabling the democratization of artificial intelligence in ophthalmology

Recent findings

Open datasets, efficient labeling techniques, code-free AutoML and cloud based platforms for deployment are resources which enable clinicians with scarce resources to drive their own AI projects.

Summary

Clinicians are the use-case experts who are best suited to drive AI projects tackling patient-relevant outcome measures. Taken together, open datasets, efficient labeling techniques, code-free AutoML and cloud platforms break the barriers for clinician-driven AI. As AI becomes increasingly democratized through such tools, clinicians and patients stand to benefit greatly.

Keywords

AI, Machine learning, artificial intelligence, labeling, cloud, democratization, data science, deep learning

Introduction

The authors strongly believe that relevant and clinically applicable deep learning research is best led by the medical use-case experts - clinicians. Based on their evolving cumulative clinical experience, clinicians are most attuned to the problem points and patient-relevant endpoints for which to optimize the algorithms of interest. Deep learning models should be designed in a use-case first approach, that is, by first identifying a use-case which addresses a real-world problem, next finding a dataset which is representative of that problem and clinical context, then applying reproducible and clinically-relevant labels, and finally training a deep learning model to predict a variable of interest for clinical care. For truly impactful work which affects clinical outcomes, clinicians are instrumental in each step of the aforementioned process.

However, full engagement of the clinical community with artificial intelligence (AI) research has significant challenges. Firstly, the development and validation of such models typically requires large volumes of data, far exceeding those required for traditional clinical prediction models[1]. However, a major enabler for AI research, particularly in ophthalmic imaging, are increasingly provisioned public datasets although the extent to which this may represent the target population and disease must be explored[2]. Alternatively, healthcare professionals may look to their own institution for local datasets, cognizant that labeling of such data may be time-consuming, and associated with significant interobserver variability. Moreover, the foreseeable information bias characteristic of electronic health records and real-world data may limit their utility as labels. Secondly, the training of complex modelling approaches, such as deep learning, require computer hardware not typically found in healthcare institutions. Changes in policy seeking to modernize the informatics infrastructure of hospitals as well as recent proposals to migrate healthcare data onto cloud systems may however provide much-needed subsistence for incorporating cloud-based Graphics Processing Units (GPUs) for model development[3,4]. Thirdly, expertise in data science and deep learning is often scarce in the clinical environment, especially among units without an academic or university affiliation. Solutions to this challenge include interdisciplinary cross-institution collaboration but also the rapidly evolving field of code-free automated machine learning (AutoML) platforms[5–7].

In this piece, we tackle the efforts and recent developments on these fronts, which are collectively empowering healthcare professionals, democratizing AI in ophthalmology, and enabling *clinician-driven* AI.

Review

Open Datasets

One of the fundamental tools enabling clinician involvement in AI research and development is the increasing availability of health data. As local datasets consisting of patient data are often relatively small, and entail barriers related to ethical approvals, open public datasets may be leveraged to design AI models. However, datasets must be carefully evaluated and scrutinised prior to being made available for use [8,9]. This step in the AI-flow requires great consideration, as the input data used to train machine learning (ML) algorithms will not only determine the outputs generated, but also the applicability and generalisability of these models in the clinical setting. In this section, we will discuss the various elements and characteristics of datasets which will allow them to truly act as a useful tool for *clinician-driven* AI.

Accessibility and Usability

To leverage available deidentified datasets for AI modeling, they need to be openly accessible, which our group has previously defined as being downloadable without fees, licensing agreements or ethical approvals [2]. We identified over 94 open access ophthalmic imaging datasets consisting of various modalities and ophthalmic diseases. In contrast to open datasets, regulated access datasets such as the UK Biobank and the AREDS datasets have strict requirements in place, which result in barriers for clinician-led development of AI models [10,11]. It is often difficult for independent healthcare professionals, especially those without affiliations to research organisations or big health institutions to fulfil the various access requirements.

While open access datasets are key in the democratization process of AI for clinicians, this data needs to be made available to the end user in a coherent and usable format. While most clinicians are familiar with image file formats such as png or jpeg, less are well-versed with formats such as MATLAB, various raw image formats, and compressed zip folders. Clinicians may be deterred by this, particularly those without access to high speed internet, resulting in long download times or failed downloads of large datasets.

Awareness is another significant barrier. While popular datasets such as Messidor in ophthalmology are well-known to the research community, there exist many lesser-known large datasets [2,12]. Having a central and up-to-date repository would not only provide greater visibility of available, it would also save time by not having to conduct literature or bibliographic searches to find relevant up-to-date datasets. One such tool recently made available to clinicians is the Google Datasets search engine, which can be used to identify up-to-date specific datasets through a user-friendly interface by indexing those published as part of academic papers and hosted on well-known dataset repositories such as Kaggle [13]. In a further effort to organize ophthalmology specific datasets for research use, our group has created eyedatasets.com, which contains links and descriptions of all known open-access ophthalmic datasets[14].

Metadata Reporting

Reporting of associated demographics and metadata is necessary for full transparency and data interrogation. This is particularly relevant to investigate data biases such as overrepresentation or underrepresenation of a particular ethnicity, age group or gender, enabling scrutiny of ML algorithms for demographic subgroup performance. Without full reporting, models trained using these datasets may have unpredictable performance, and possibly unintended consequences for subgroups of the general population [15,16]. Complete metadata reporting would also allow clinicians to combine images across smaller datasets to develop larger compiled datasets. This is especially advantageous for uncommon ophthalmic diseases and imaging techniques, where large datasets from a single institution or research group may be difficult to obtain.

Relevance and Representation

In a recent review, the concept of health data poverty was highlighted, where there was unequal representation of the global population across ophthalmic datasets [17]. This is an issue that needs to be addressed. If models are trained using datasets which are biased to a certain demographic, they may not perform as well if a clinician’s practice has individuals from an underrepresented category. First, this has the great potential to exacerbate health disparities. Second, clinicians may be less incentivised to drive AI development if their patient community does not benefit due to lack of available representative datasets.

Open Datasets: A Resource for AI Democratization

Despite the aforementioned considerations, existing open health datasets are one of the key drivers of *clinician-driven* AI, as they allow clinicians to leverage large amounts of data for ML development. Through the use of open datasets, clinicians may curtail obtaining ethical approvals for collating and de-identifying large volumes of their own institution's patient data to curate a dataset. When proper ethical approvals are obtained, and local data is available, open datasets may also be utilised in ML algorithm testing as additional external validation sets. In healthcare organizations where patient data has not yet been digitised, open datasets are a valuable asset to enable clinciain familiarity with AI research and development.

Efficient Labeling Techniques

Traditional labeling process

In the context of healthcare, most ML algorithms are designed with a supervised learning approach, which relies on labeled data for model training. Accurate and reproducible labels are required to achieve high model performance, and the following points should be kept in mind during the labeling process [18]. When selecting the labeling workforce, it is important to consider that experts have unique strengths at classifying varying types of data [19]. Moreover, the choice of the number of graders depends on factors such as the subjectivity of the task, the time and the resources available for each project [20,21]. In addition to qualified graders, it's desirable to have a grading guideline containing detailed and specific definitions of the ground truth, as well as some examples to illustrate subjective grading schemes [22]. Lastly, it is important to carefully decide how to resolve the disagreements that naturally arise when working with multiple graders. The choice between methods such as majority vote, arbitration by a more senior grader, or adjudication should aim to find a balance between the reliability, time and costs involved with each method [23,24].

Challenges in Data Labeling

Manually labeling data is highly costly and time-consuming, and the scarcity of experts to handle the data makes the annotation step a bottleneck of the model training process [25–29]. In addition, medical tasks often involve great subjectivity, grader variability and noise in training labels. Techniques such as adjudication by multiple graders can improve label quality and provide a rigorous reference standard for model training, but the high time costs prevent this method from being used more often [23,30]. These barriers are more pronounced in low-income countries and small research groups, which hinders the democratization of AI [6].

Deep Learning Techniques Utilizing Fewer Labels

Many studies have been carried out to decrease the resource requirements involved in the traditional labeling process (i.e., manual labeling of data) [27,30–32]. Semi-supervised learning (SSL) is a deep learning method that makes use of both labeled and unlabeled data, decreasing the need for high quality labeled instances, and achieving superior accuracy when compared to unsupervised methods, which use only unlabeled data for model training [25,26,33]. For example, in self-training, an SSL technique, a model (i.e., teacher model) is firstly trained on a small, labeled set and used to predict pseudolabels on the large, unlabeled set. A second model (i.e., student model) is then trained on the whole dataset, containing labeled and pseudolabeled data [26,34,35]. This process may be iterated (i.e., putting the student back as a teacher) until the desired performance is reached [36].

Active learning (AL) is another deep learning technique which allows efficient use of unlabeled data. In this method, the most informative instances for model training are selected for labeling, aiming to reduce cost, while maintaining an adequate model performance [26,35,37,38]. AL has been recently evaluated in the context of diabetic retinopathy, in which large amounts of unlabeled data generated by screening programs represent a use case for this technique [39,40]. This and other alternatives to reduce labeling costs (e.g., transfer learning and self-supervised learning) represent enabling techniques for research by clinicians with limited resources. Further work is required to determine an appropriate balance between the contributions of humans and automated systems in the ML pipeline [41].

AutoML Platforms

Automated machine learning (AutoML) refers to the process of automating some or all parts of the MLworkflow, obviating the need for coding expertise [6]. Currently, several AutoML platforms are available, offering a wide range of services that include data cleaning, model selection, hyperparameter optimization and, in some cases, labeling tools [5]. The Google Cloud AutoML Vision platform, for example, offers the “Data Labelling Service”, through which users can submit instructions containing detailed definitions of the ground truth and receive the data annotated by human labelers [42]. The Microsoft Azure Custom Vision platform offers a tool called “Smart Labeler”, which resembles the self-training technique. Firstly, the user uploads the data, labels part of the training set and starts the automated training process. Once the training is completed, the platform uses the latest trained iteration of the model to predict the labels of unlabeled images [43]. Although these tools have the potential to facilitate the annotation step, their feasibility in clinical data has yet to be evaluated.

Code-free AutoML

AutoML Platforms

A resource which curtails the need for coding-expertise enabling *clinician-driven* ML is AutoML. This comprises a set of tools, available from multiple vendors (Amazon, Apple, Clarifai, Google, MedicMind, Microsoft and others), which automate the ML process[5]. These publicly accessible platforms enable a clinician-friendly approach where the AutoML provides a user interface to upload and collate data, assign labels, and train a deep learning model without coding. A number of services run in the cloud (Amazon, Clarifai, Google, Microsoft, MedicMind), provisioning the requisite graphical processing units or tensor processing units, and thereby removing the clinician-facing obstacle of providing their own expensive hardware to train deep learning models. The platforms typically train the models automatically, and perform steps such as selection of optimal network architectures, pre-processing methods and hyperparameter optimization.

Platform Performance and Features

When the model is finished training, it may be deployed to the cloud and made available for evaluation and prediction. Some platforms (Google, Microsoft) allow download of the resulting model, for use on mobile devices as an “edge model” or for local evaluation and use. Our group has performed a systematic performance and featureset review of available code-free platforms on ophthalmic imaging of various modalities[5]. Three leading platforms (Amazon, Google, Microsoft), demonstrated uniformly high ML model performance, often higher than bespoke coded ML models in the literature which utilized the same datasets for training. However, only the Google platform contained all features recognized as crucial for clinically useful ML model evaluation and explainability such as confusion matrices, precision-recall curves, saliency maps, and providing threshold adjustments[44–47]. The Google platform provided all of these features, the Amazon platform provided none, and the Microsoft platform only provided threshold adjustment. External validation of ML models is crucial before clinical use[48]. Crucially, of these leading three platforms, only Google provided scalable batch prediction capability, which enables external validation by rapidly producing predictions on large datasets.

An important consideration of the available AutoML platforms is cost. While clinicians may save money by avoiding purchasing expensive local hardware to train ML models, the cloud-based platforms have a limited free tier before requiring payment for their compute services. Thus, performance per dollar will become a key metrics for those researchers and clinicians who are budget constrained. As personal computing power becomes more powerful, these costs will inevitably decrease, allowing for further democratization. As these platforms mature, the ML process will become increasingly commoditized, and clinicians will become more versed in utilizing the available tools.

AutoML Enabling Democratization

Code-free AutoML is a robust and increasingly available resource which democratizes ML access for *clinician-driven* AI. In addition to image classification and segmentation models, it has a wide range of uses including natural language processing, tabular data, and video data, all of which is being generated at a rapid rate in our clinics. AutoML's use-cases are broad, and models may be rapidly trained by the clinicians who are uniquely attuned to their respective bottlenecks including but not limited to patient care, clinic flow, referral management, and research study recruitment. The increased ML exposure these platforms provide clinicians will not only promote fast iteration and problem solving, but also drive exploration and knowledge of responsable AI practices[49].

Model Deployment: Ingest Data, Provide Predictions via User Interface

At a rapid pace, new ML models are emerging from the research space. Although promising, almost none of the ML models have made it to the clinic, as crucial obstacles still stand in the way of the widespread use of AI in this setting. This is demonstrated by only two ophthalmic diabetic retinopathy algorithms having received FDA approval, and only a fraction of ophthalmologists utilizing AI in their daily work[50,51]. To take an AI model from research to the clinic, the model must be packaged and deployed. This involves creating an end-to-end solution which goes from a medical scan ingestion to the delivery of the AI-derived insights to the doctor. The high level steps to create this solution include: utilizing quality controlled image data, feeding the data to the model to produce predictions, delivering the predictions to the user through a tailored interface, and finally, obtaining regulatory approval.

Real-time Data Ingestion

ML model training starts with the creation of a dataset. Specifically in medical imaging, this involves either using an existing public dataset or extracting a dataset manually from the local hospital imaging file systems. In contrast, during model *deployment*, the data will originate from the local hospital imaging file systems. This implies that one must connect with the local file-system that stores the images, identify the scans of interest, ingest those scans in real-time, and convert them to a standard format which is usable by the model.

As every medical imaging vendor uses varying storage systems and metadata structures, connecting to the clinical image file-system may be challenging. Moreover, the images are often stored in proprietary file formats which may only be viewed via vendor-specific software, and in most cases no batch converter is provided to convert or extract images to standard file formats. As a result, to ingest data for a model in real-time, file format converters are necessary for every scanning device and vendor, alternatively, the images may be stored in a standard format, similar to the DICOM standard in radiology. Recently, the American Academy of Ophthalmology (AAO) stated that ophthalmic imaging device manufacturers should standardize image formats and that “adherence to standardization would advance the needs and interests of ophthalmologists, their patients, and the quality of clinical care by promoting interoperability, enabling the creation of comprehensive datasets for research and big data analyses, and developing algorithms for ML and artificial intelligence”[52].

Generating Model Predictions: Local, Cloud, or Smartphone

Today, the use and impact of models is not just limited by technical hardware capabilities, but also by the operational processes to develop, share, and deploy ML models. Conventions for ML model standardization are required for the launch of real-world AI in healthcare. Common conventions (as the MLCommons initiative has proposed in MLCube) can ensure models behave as interchangeable building blocks, which may be easily shared for experimentation between researchers and hospitals[53]. This will eventually enable enhanced collaboration, and increasingly robust workflows.

Before model deployment, a ML developer should decide whether it will be run locally, in the cloud, or on a smartphone. Previously, hospitals focused on having their IT infrastructure locally 'on-premise', as this allowed them to set up and control the full IT setting entirely independently, initially coming with the benefit of privacy and security. However, as those systems grow in complexity over the years, adding new features, building upon old software, this can result in highly complex systems which are difficult to maintain, with increasing outages and cybersecurity risks. The ability to build ML solutions on local infrastructure is frequently hampered by both limitations in access to compute resources, and restrictions on the installation of ML software.

Hospitals are increasingly seeking to modernize their informatics infrastructure, as well as migrating healthcare data onto cloud systems. Storing and hosting databases on cloud systems (e.g. Microsoft Azure, Amazon AWS, Google Cloud) allows users to benefit from the expertise of major technology companies in security, scaling, and ensuring outage-proof stable infrastructure. Additionally, moving to the cloud allows for easier integration with purpose built ML tools in the cloud such as AutoML. Further, it enables efficient scaling of compute resources such as cloud-based graphics processing units for model development or deployment on-demand. Cloud companies are also focusing on 'serverless' services, meaning the developer does not need to manage and set up their own servers, scale resources in case of increased demand, create backups, or guarantee security. This allows developers to truly focus on building ML solutions instead of thinking about servers, security, and backup infrastructure.

Beside deployment on local and cloud infrastructure, ML models may also be hosted on smartphone devices. Both TensorFlow and PyTorch, the two most commonly used ML frameworks, are expanding their lightweight libraries to deploy models on mobile and embedded devices (TensorFlow Lite, PyTorch mobile respectively)[54,55]. Not only are smartphone chips surprisingly suited for producing ML predictions, they also allow physicians to make predictions completely locally without the potential privacy issues of having to send data to the cloud as in a cloud-based setup. In 2020, the first open dataset of smartphone pictures of medical scans was created[56]. A follow-up paper concluded that although the models, used for classifying chest X-ray, had a drop in performance when applied to photos of chest X-rays, some models still performed comparably to radiologists [57].

User Interface

The interface software to present ML insights should be adapted to the clinical workflow[58]. In other industries, custom software interfaces are increasingly built by “citizen developers”. Although their primary job function is outside the field of developing or ML analytics, these citizen developers create new software applications primarily using no-code initiatives. No-code initiatives for general web and mobile applications allow non-technical users to build web and mobile apps without coding expertise. The recent developments in no-code and low-code initiatives may also facilitate the emergence of citizen developers in medicine. The creation of interfaces built by and for physicians will result in interfaces which fit within the physician workflow with integrated and helpful interactions. Eventually, these interfaces will allow physicians to not only build models with AutoML, but also launch their own models for use in their own clinic.

Conclusion

Through the proliferation of open datasets, efficient labeling techniques, code-free AutoML, and cloud-based model deployment frameworks, the barriers for *clinician-driven* ML are significantly decreasing. Gartner research defines a citizen data scientist as a “person who creates or generates models that use advanced diagnostic analytics [...], but whose primary job function is outside the field of statistics and analytics”[59]. We believe that clinicians, as the use-case experts, are best suited to be the clinical citizen data scientists who design ML models for their own use to better patient care. Utilizing the highlighted resources which enable the democratization of ML, clinicians who were previously interested in applying ML, yet hesitant due to the perceived hurdles, can now lead the progress of *clinician-driven* AI.

Key Points

- *Clinician-driven* AI projects are now possible, even by researchers and clinicians without coding experience through publicly available tools
- Open datasets, AutoML, efficient labeling techniques, and cloud platforms enable machine learning model design with scarce resources
- *Clinician-driven* AI has the potential to address the AI chasm, enabling use-case driven projects for relevant endpoints with higher likelihood of actual implementation

Acknowledgements

Acknowledgements: None

Financial Support and Funding:

EK: Springboard Grant Moorfields Eye Charity R190016A MBG: Lemann Foundation; Vision Institute (IPEPO); CAPES Foundation, Ministry of Education of Brazil, Brasília, DF, Brazil

SK: N/A

RS: N/A

SW: N/A

PAK: Moorfields Eye Charity Career Development Award (R190028A) and a UK Research & Innovation Future Leaders Fellowship (MR/T019050/1)

Conflicts of Interest:

EK is a former employee of Google Health, he has acted as a consultant for Genentech and is an equity owner in Reti Health. PAK has acted as a consultant for DeepMind, Roche, Novartis, Apellis, and BitFount and is an equity owner in Big Picture Medical. He has received speaker fees from Heidelberg Engineering, Topcon, Allergan, and Bayer. The remaining authors report no conflicts of interest.

References

1. Najafabadi MM, Villanustre F, Khoshgoftaar TM, Seliya N, Wald R, Muharemagic E. Deep learning applications and challenges in big data analytics. Journal of Big Data. 2015 Feb 24;2(1):1–21.

2. **Khan SM, Liu X, Nath S, Korot E, Faes L, Wagner SK, et al. A global review of publicly available datasets for ophthalmological imaging: barriers to access, usability, and generalisability. Lancet Digit Health. 2021 Jan;3(1):e51–66.

This paper catalogues all available public ophthalmic datasets, which may be used by readers to train their own machine learning models.

3. Organization WH, Others. Draft global strategy on digital health 2020--2024. World Health Organization, ed Geneva. 2019;

4. *NHS and social care data: off-shoring and the use of public cloud services [Internet]. [cited 2021 Jun 4]. Available from: https://digital.nhs.uk/data-and-information/looking-after-information/data-security-and-information-governance/nhs-and-social-care-data-off-shoring-and-the-use-of-public-cloud-services

This paper is one of the first guidances of public cloud services, which will lead to increased data availability for healthcare.

5. **Korot E, Guan Z, Ferraz D, Wagner SK, Zhang G, Liu X, et al. Code-free deep learning for multi-modality medical image classification. Nature Machine Intelligence [Internet]. 2021 Mar 1; Available from: https://doi.org/10.1038/s42256-021-00305-2

This paper describes a systematic review of automated machine learning platforms. It details performance and featureset evaluations of six publicly available platforms for code-free machine learning.

6. Faes L, Wagner SK, Fu DJ, Liu X, Korot E, Ledsam JR, et al. Automated deep learning design for medical image classification by health-care professionals with no coding experience: a feasibility study. The Lancet Digital Health. 2019 Sep 1;1(5):e232–42.

7. *Korot E, Pontikos N, Liu X, Wagner SK, Faes L, Huemer J, et al. Predicting sex from retinal fundus photographs using automated deep learning. Sci Rep. 2021 May 13;11(1):10286.

This paper demonstrates that automated deep learning performance can rival that of bespoke models. It also shows high performance of deep learning on a task which humans cannot do, predict sex from retinal fundus photographs.

8. Altman ER. Big Data and Democratization [Editorial]. IEEE Micro. 2014 Jul;34(4):2–3.

9. Fahey S. The Democratization of Big Data. J Nat’l Sec L & Pol'y. 2014;7:325.

10. dbGaP Study [Internet]. [cited 2021 Jun 3]. Available from: https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000001.v3.p1

11. Chua SYL, Thomas D, Allen N, Lotery A, Desai P, Patel P, et al. Cohort profile: design and methods in the eye and vision consortium of UK Biobank. BMJ Open. 2019 Feb 21;9(2):e025077.

12. Decencière E, Zhang X, Cazuguel G, Lay B, Cochener B, Trone C, et al. Feedback on a publicly distributed image database: The Messidor database. Image Anal Stereol. 2014 Aug 26;33(3):231.

13. *Dataset Search [Internet]. [cited 2021 Jun 3]. Available from: https://datasetsearch.research.google.com/help

Dataset search is a highly useful tool enabling researchers to find new datasets for their machine learning projects as they become available.

14. Ophthalmology Imaging Datasets [Internet]. [cited 2021 Jun 4]. Available from: https://www.eyedatasets.com/

15. *Futoma J, Simons M, Panch T, Doshi-Velez F, Celi LA. The myth of generalisability in clinical research and machine learning in health care. Lancet Digit Health. 2020 Sep;2(9):e489–92.

This paper details the importance of external validation for robust an equitable machine learning.

16. Vollmer S, Mateen BA, Bohner G, Király FJ, Ghani R, Jonsson P, et al. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. BMJ. 2020 Mar 20;368:l6927.

17. Ibrahim H, Liu X, Zariffa N, Morris AD, Denniston AK. Health data poverty: an assailable barrier to equitable digital health care. Lancet Digit Health. 2021 Apr;3(4):e260–5.

18. Willemink MJ, Koszek WA, Hardell C, Wu J, Fleischmann D, Harvey H, et al. Preparing Medical Imaging Data for Machine Learning. Radiology. 2020 Apr;295(1):4–15.

19. Guan MY, Gulshan V, Dai AM, Hinton GE. Who Said What: Modeling Individual Labelers Improves Classification [Internet]. arXiv [cs.LG]. 2017. Available from: http://arxiv.org/abs/1703.08774

20. Langlotz CP, Allen B, Erickson BJ, Kalpathy-Cramer J, Bigelow K, Cook TS, et al. A Roadmap for Foundational Research on Artificial Intelligence in Medical Imaging: From the 2018 NIH/RSNA/ACR/The Academy Workshop. Radiology. 2019 Jun;291(3):781–91.

21. Rizos G, Schuller BW. Average Jane, Where Art Thou? – Recent Avenues in Efficient Machine Learning Under Subjectivity Uncertainty. In: Information Processing and Management of Uncertainty in Knowledge-Based Systems. Springer International Publishing; 2020. p. 42–55.

22. Mongan J, Moy L, Kahn CE. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers. Radiology: Artificial Intelligence. 2020 Mar 1;2(2):e200029.

23. Krause J, Gulshan V, Rahimy E, Karth P, Widner K, Corrado GS, et al. Grader Variability and the Importance of Reference Standards for Evaluating Machine Learning Models for Diabetic Retinopathy. Ophthalmology. 2018 Aug;125(8):1264–72.

24. Schaekermann M, Hammel N, Terry M, Ali TK, Liu Y, Basham B, et al. Remote Tool-Based Adjudication for Grading Diabetic Retinopathy. Transl Vis Sci Technol. 2019 Nov;8(6):40.

25. Gu L, Zhang X, You S, Zhao S, Liu Z, Harada T. Semi-Supervised Learning in Medical Images Through Graph-Embedded Random Forest. Front Neuroinform. 2020 Nov 10;14:601829.

26. Enguehard J, O'Halloran P, Gholipour A. Semi Supervised Learning with Deep Embedded Clustering for Image Classification and Segmentation. IEEE Access. 2019 Jan 9;7:11093–104.

27. Xue Y, Hauskrecht M. Efficient Learning of Classification Models from Soft-label Information by Binning and Ranking.pdf. Proc Int Fla AI Res Soc Conf. 2017. p. 164–9.

28. Bellemo V, Lim ZW, Lim G, Nguyen QD, Xie Y, Yip MYT, et al. Artificial intelligence using deep learning to screen for referable and vision-threatening diabetic retinopathy in Africa: a clinical validation study. The Lancet Digital Health. 2019 May 1;1(1):e35–44.

29. Korot E, Wood E, Weiner A, Sim DA, Trese M. A renaissance of teleophthalmology through artificial intelligence. Eye . 2019 Jun 1;33(6):861–3.

30. Hsu J, Phene S, Mitani A, Luo J, Hammel N, Krause J, et al. Improving Medical Annotation Quality to Decrease Labeling Burden Using Stratified Noisy Cross-Validation [Internet]. arXiv [cs.LG]. 2020. Available from: http://arxiv.org/abs/2009.10858

31. Fries JA, Varma P, Chen VS, Xiao K, Tejeda H, Saha P, et al. Weakly supervised classification of aortic valve malformations using unlabeled cardiac MRI sequences. Nat Commun. 2019 Jul 15;10(1):3111.

32. Sucholutsky I, Schonlau M. “Less Than One”-Shot Learning: Learning N Classes From M<N Samples [Internet]. arXiv [cs.LG]. 2020. Available from: http://arxiv.org/abs/2009.08449

33. Azary H, Abdoos M. A Semi-Supervised Method for Tumor Segmentation in Mammogram Images. J Med Signals Sens. 2020 Jan;10(1):12–8.

34. Nartey OT, Yang G, Wu J, Asare SK. Semi-Supervised Learning for Fine-Grained Classification With Self-Training. IEEE Access. 2020;8:2109–21.

35. Li M, Zhou Z-H. SETRED: Self-training with Editing [Internet]. Advances in Knowledge Discovery and Data Mining. 2005. p. 611–21. Available from: http://dx.doi.org/10.1007/11430919_71

36. *Xie Q, Luong M-T, Hovy E, Le QV. Self-Training With Noisy Student Improves ImageNet Classification [Internet]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020. Available from: http://dx.doi.org/10.1109/cvpr42600.2020.01070

This paper details an efficient labeling technique which enables leveraging large unlabeled datasets for improved model performance.

37. Fazakis N, Kanas VG, Aridas CK, Karlos S, Kotsiantis S. Combination of Active Learning and Semi-Supervised Learning under a Self-Training Scheme. Entropy [Internet]. 2019 Oct [cited 2020 Dec 22];21(10). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7514320/

38. Ouali Y, Hudelot C, Tami M. An Overview of Deep Semi-Supervised Learning [Internet]. arXiv [cs.LG]. 2020. Available from: http://arxiv.org/abs/2006.05278

39. Zhang Y, An M. An Active Learning Classifier for Further Reducing Diabetic Retinopathy Screening System Cost. Comput Math Methods Med. 2016 Aug 29;2016:4345936.

40. Ahsan MA, Qayyum A, Qadir J, Razi A. An Active Learning Method for Diabetic Retinopathy Classification with Uncertainty Quantification [Internet]. arXiv [cs.LG]. 2020. Available from: http://arxiv.org/abs/2012.13325

41. Budd S, Robinson EC, Kainz B. A survey on active learning and human-in-the-loop deep learning for medical image analysis. Med Image Anal. 2021 Apr 9;71:102062.

42. Requesting data labeling [Internet]. [cited 2021 Jun 4]. Available from: https://cloud.google.com/vertex-ai/docs/datasets/data-labeling-job?hl=fa

43. Farley P. Label images faster with Smart Labeler [Internet]. [cited 2021 Jun 4]. Available from: https://docs.microsoft.com/en-us/azure/cognitive-services/custom-vision-service/suggested-tags

44. Choo J, Liu S. Visual Analytics for Explainable Deep Learning. IEEE Comput Graph Appl. 2018;38(4):84–92.

45. Ribeiro MT, Singh S, Guestrin C. “Why Should I Trust You?”: Explaining the Predictions of Any Classifier [Internet]. arXiv [cs.LG]. 2016. Available from: http://arxiv.org/abs/1602.04938

46. Vickers AJ, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. BMC Med Inform Decis Mak. 2008 Nov 26;8:53.

47. Shah NH, Milstein A, Bagley PhD SC. Making Machine Learning Models Clinically Useful. JAMA [Internet]. 2019 Aug 8; Available from: http://dx.doi.org/10.1001/jama.2019.10306

48. Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol. 2016 Jan;69:245–7.

49. *Keane PA, Topol EJ. AI-facilitated health care requires education of clinicians. Lancet. 2021 Apr 3;397(10281):1254.

This paper describes the importance of raising clinician AI literacy.

50. Eyenuk Announces FDA Clearance for EyeArt Autonomous AI System for Diabetic Retinopathy Screening - Eyenuk, Inc. ~ Artificial Intelligence Eye Screening [Internet]. [cited 2021 Jun 4]. Available from: https://www.eyenuk.com/us-en/articles/diabetic-retinopathy/eyenuk-announces-eyeart-fda-clearance/

51. Office of the Commissioner. FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems [Internet]. 2018 [cited 2021 Jun 4]. Available from: https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-based-device-detect-certain-diabetes-related-eye

52. Lee AY, Campbell JP, Hwang TS, Lum F, Chew EY, American Academy of Ophthalmology. Recommendations for Standardization of Images in Ophthalmology. Ophthalmology [Internet]. 2021 Apr 5; Available from: http://dx.doi.org/10.1016/j.ophtha.2021.03.003

53. MLCube™ [Internet]. [cited 2021 Jun 4]. Available from: https://mlcommons.org/en/mlcube/

54. TensorFlow Lite [Internet]. [cited 2021 Jun 4]. Available from: https://www.tensorflow.org/lite

55. PyTorch Mobile [Internet]. [cited 2021 Jun 4]. Available from: https://pytorch.org/mobile/home/

56. CheXphoto: A Dataset of Naturally and Synthetically Perturbed Chest X-Rays and Competition for Automated X-Ray Interpretation [Internet]. [cited 2021 Jun 4]. Available from: https://stanfordmlgroup.github.io/competitions/chexphoto/

57. Rajpurkar P, Joshi A, Pareek A, Irvin J, Ng AY, Lungren M. CheXphotogenic: Generalization of Deep Learning Models for Chest X-ray Interpretation to Photos of Chest X-rays [Internet]. arXiv [eess.IV]. 2020. Available from: http://arxiv.org/abs/2011.06129

58. Beede E, Baylor E, Hersch F, Iurchenko A, Wilcox L, Ruamviboonsuk P, et al. A Human-Centered Evaluation of a Deep Learning System Deployed in Clinics for the Detection of Diabetic Retinopathy. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. New York, NY, USA: Association for Computing Machinery; 2020. p. 1–12. (CHI '20).

59. Citizen Data Science Augments Data Discovery and Simplifies Data Science [Internet]. [cited 2021 Jun 4]. Available from: https://www.gartner.com/en/documents/3534848