Artificial Intelligence in PET: An Industry Perspective

Arkadiusz Sitek, PhD[a,*], Sangtae Ahn, PhD[b], Evren Asma, PhD[c], Adam Chandler, PhD[d], Alvin Ihsani, PhD[e], Sven Prevrhal, PhD[f], Arman Rahmim, PhD, DABSNM[g,h], Babak Saboury, MD, MPH, DABR, DABNM[i,j,k], Kris Thielemans, PhD[l,m]

KEYWORDS

• Industry • AI • List-mode • Ecosystem • Workflow • PET

KEY POINTS

- Industry faces unique challenges to bring artificial intelligence (AI) to positron emission tomography (PET) clinical workflows.
- There are new AI ecosystems created to facilitate the use of AI in clinics.
- New computing ecosystems can include reconstructions of vendor neutral format raw PET list-mode data.
- Custom workflows including image reconstructions and list-mode data processing can be used in new AI ecosystems.

GLOSSARY

Artificial intelligence, convolutional neural networks, Positron Emission Tomography, Radiology, AI ecosystem, AI workflows, federated learning, industry perspective, data acquisition, list-mode data, standardization, cost, data access, robustness, underspecification of AI model, clinical value, regulations, AI failures, adversarial attacks, uncertainty estimation, explainability, decision making, decision support, human-machine decision making, liability, custom data processing, adoption of AI, trust in AI recommendations, PET list-mode standardization, standardized image reconstruction.

INTRODUCTION

The recent popularity of *artificial intelligence* (AI) heralded as a game-changing technology has generated high hopes for breakthrough advancements and changes across the entire health care industry. The specific area of clinical positron emission tomography (PET) imaging is no exception. In this work, we provide an industry perspective on specific opportunities and challenges for PET arising by the emergence of AI and deep learning (DL) methods.

DL[1] is a machine learning technique which uses deep neural networks to create a variety of models which can process raw data. In recent years, DL

[a] Sano Centre for Computational Medicine, Nawojki 11 Street, Kraków 30-072, Poland; [b] GE Research, 1 Research Circle KWC-1310C, Niskayuna, NY 12309, USA; [c] Canon Medical Research, 706 N Deerpath Drive, Vernon Hills, IL 60061, USA; [d] Global Scientific Collaborations Group, United Imaging Healthcare, America, 9230 Kirby Drive, Houston, TX 77054, USA; [e] NVIDIA, 2 Technology Park Drive, Westford, MA 01886, USA; [f] Philips Research Europe, Röntgenstr. 22, Hamburg 22335, Germany; [g] Department of Radiology, University of British Columbia, BC Cancer, BC Cancer Research Institute, 675 West 10th Avenue, Office 6-112, Vancouver, British Columbia V5Z 1L3, Canada; [h] Department of Physics, University of British Columbia, BC Cancer, BC Cancer Research Institute, 675 West 10th Avenue, Office 6-112, Vancouver, British Columbia V5Z 1L3, Canada; [i] Department of Radiology and Imaging Sciences, Clinical Center, National Institutes of Health, 9000 Rockville Pike, Bethesda, MD 20892, USA; [j] Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County, Baltimore, MD, USA; [k] Department of Radiology, Hospital of the University of Pennsylvania, 3400 Spruce Street, Philadelphia, PA 19104, USA; [l] Institute of Nuclear Medicine, University College London, UCL Hospital Tower 5, 235 Euston Road, London NW1 2BU, UK; [m] Algorithms and Software Consulting Ltd, 10 Laneway, London SW15 5HX, UK
* Corresponding author.
E-mail address: a.sitek@sanoscience.org

PET Clin 16 (2021) 483–492
https://doi.org/10.1016/j.cpet.2021.06.006

has demonstrated significantly promising results for several PET applications, including segmentation, reconstruction, outcome modeling, decision support, and so forth.[2–8]

In this work, a *manufacturer* is defined as an industry member manufacturing PET scanners, and a *vendor* as an industry member providing AI and other processing software solutions. These two groups are not exclusive. In the present work, we sometimes interchangeably use the terms *deep learning* and *artificial intelligence*, although AI (and its subset machine learning) are wider fields. The recent spike of interest in AI is due to increased popularity of DL, especially the use of convolutional neural networks (CNNs), which is why we use this convention in this article.

The article is organized as follows. Section Challenges for commercialization identifies selected challenges in the adoption of AI from the industry point of view. They are general and not PET specific. The goal of this section is not to discuss potential solutions to those challenges but rather paint a perspective on specific challenges from the industry point of view. In section Looking into the future of AI in PET, specific applications of AI in PET are discussed in more detail. In particular, a concept of reconstruction of the list-mode (LM) data on demand is combined with AI algorithms and presented.

CHALLENGES FOR COMMERCIALIZATION

One of the main concerns for industry is to release reliable, extensively tested and validated products that impact disease and patient management. For the purpose of this article, we define a reliable product as working as intended and within a set of predefined specifications. Equally important, the product should demonstrate clinical utility. In this section, we discuss some of the major concerns and obstacles the industry has to overcome.

Development and Clinical Evidence

Access to data

Obtaining large amounts of data to develop AI products is challenging, and often ownership of the data is not with the industry. The need for obtaining sufficient amounts of data for training, which encompass all expected variations in the data, that is, population-based variations (both locally and geographically), body locations, disease state variations (including normal/nondisease cases), and so forth, adds further challenges. *Federated learning* (FL) is an approach that may at least partially alleviate the issue. In FL, AI models are trained based on data that never leave the medical institutions[9] and therefore data security and privacy are much less of a concern. The paradigm of FL is being widely explored (eg, the work on FL from the London Medical Imaging & AI Center for Value-Based Healthcare[10,11]).

Ground truth

In some applications of AI such as supervised learning, obtaining ground truth will present a great challenge. Ground truth can be obtained from an independent measurement (eg, biopsy, postmortem analysis), clinical outcomes (death, morbidity), or previous diagnoses (eg, radiology reports), or new reads or annotations can be used. Ideally, the data sets should be large, but new reads and annotations make data preparation a lengthy and expensive process.

Robustness

Of particular, commercial interest is a reliable, regulatory cleared product that performs according to specifications regardless of geographic location, patient mix, and local preferences and guidelines. Unfortunately, AI algorithms can generalize poorly and are dependent on the data sets used to train and test the algorithms. An AI algorithm may produce unreliable results if characteristics of the input deviate from the training data. This has critical consequences. It is acceptable to publish an AI algorithm tested on homogeneous data (eg, from a single or small number of institutions using well-defined study inclusion criteria) as long as those limitations are transparently disclosed in the publication. However, a commercially available product ought to be applied to real life data that may be more diverse and complex than single-center study data, which may render certain limitations of an algorithm as nonacceptable. In general radiology, there are many large publicly available data sets which can be used to test generalization of developed AI algorithms. Unfortunately, there are few such sets available that include PET data, making the development of AI algorithms for PET more difficult.

Underspecification

Another obstacle to generalization of AI is a recently documented problem of *underspecification*.[12] This term denotes the problem that if we train the same model a number of times with slightly different initial weights on the same large data set and achieve similar performance on the test set, there is no guarantee that those models will perform the same in the real world. This is a very difficult problem to tackle and extremely important from the industry point of view as the real-world performance is what matters. When many models are trained on the same set of data with random initial weights and applied to a certain unseen real-world scenario, some models may work and some other models may not. At the phase of model development, it is difficult to tell which of those models will work and which ones will not. Testing models with diverse real-world data will alleviate the problem although not entirely. Therefore, we emphasize the importance of postmarket surveillance after algorithm deployment which becomes even more important than classical (not DL) offerings.

Clinical value

When commercializing AI algorithms, there is a need to demonstrate that the product provides clinical value and evidence that supports the intended use. To generate such data that can be used as evidence for potential regulatory claims that translate into customer value, often multicenter, multireader studies are required. Here, we emphasize that often one develops an excellent technological solution to a clinical problem, but when introduced to clinical workflows, it is not widely used in routine clinical practice by clinicians. Appropriately designed external evaluation studies at clinical sites by clinicians could mitigate the problem.

Regulatory Pathways

AI's towering dependence on data exposes MedTechś regulatory and privacy challenges more than ever before: compounded by the sharp teeth that GDPR has afforded the EU, with global effects, academia and industry are only now learning to safely share massive amounts of data.

Regulatory bodies, too, increasingly demand being shown the data used to train the AI parts of software submitted for their approval. However, basing approval on the data creates the conundrum as once approved that retraining with new data would invalidate it and burden industry and administration with incessant reapproval cycles. Luckily, everybody agrees a solution is direly needed. In the United States, the FDA is working on an action plan, and the EU has just released a white paper with very similar thoughts.[13–15] Obviously, an eventual worldwide joint framework will be key for industry and data-owning individuals alike.

Return on Investment

The health care industry requires a reasonable return of investment to create or sustain a viable business. For applications of AI in PET products, investment in development should be properly justified by balancing the growth potential of the AI technology with the considerable risks. AI may require a nontraditional business model in which subscription approaches, architectures open to third parties such as marketplaces, and new ecosystems are used. It is an industry challenge to figure out why and how clients would pay for AI innovations.

Understanding AI

Explainability

Explainability is an important factor associated with the adoption of AI from the commercial point of view. In short, in DL methods, the decisions made by AI are often opaque, black box decisions. For more details on this problem, please refer to the study by Arrieta and colleagues.[16] For AI algorithms to succeed in the commercial world, the users of the algorithms have to gain trust in them. For example, a clear explanation on how the AI algorithm arrived at a certain classification can increase trust in the subsequent clinical decisions which AI recommends.

Education and trust

It is critically important to educate users about AI's capabilities and, even more importantly, its limitations. Most current applications of neural networks are some form of image denoising where very noisy images, presumably from short- or low-dose scans, are converted into images that appear less noisy. However, this does not mean that nothing is lost because of shorter or lower dose scans. Users need to understand that the quantitative lesion/region of Interest performances of their images are still governed by the statistics of the acquisition. AI can mimic longer or higher dose scans by making backgrounds smoother but cannot create the information that is lost due to shorter or lower dose scans. Nonetheless, we note that AI can improve image quality such as lesion detectability or signal-to-noise ratio by using better priors, system models, data correction, or noise models learned from data.

Another related topic is how clinicians determine the reliability of lesion standardized uptake values. They may look at how noisy a large, approximately constant region such as the liver is and decide that smoother regions indicate more reliable lesion quantitation. For typical reconstruction algorithms such as Ordered-subset expectation maximization (OSEM), this approach works reasonably well because if the noise correlation lengths are short, the single-image-noise in the liver is related to the standard deviation of a single liver voxel, which is in turn related to the standard deviations of individual lesion voxels which finally determine the standard deviation of the lesion SUV. In contrast, when the background is smoothed using

AI-based methods, this connection is lost. The single-image-noise in the liver may be greatly reduced without any significant changes to the lesion ensemble noise properties. Therefore, a clinician looking at an image denoised with neural networks should be cautious about interpreting the variability (or uncertainty) of the lesion SUV. It should also be noted that denoising could introduce an additional bias in the lesion SUV.

Combining human and AI insights

In the foreseeable future, human decision-makers will be augmented/assisted and not replaced by automated algorithms. Unavoidably there will be situations where a human opinion is different from that of an algorithm. This creates opportunities and challenges because the combination of AI and humans may create a better and more accurate decision.[17,18] However, it creates a problem on how to meaningfully combine human and AI insights. The final decision in the foreseeable future will be made by humans, and some solutions are needed to deal with disagreements. One such approach could be that AI provides explanations or examples from the past of similar images with known outcomes, which may persuade the physician. Another resolution of such conflicts could be that we teach the AI algorithm to consider the physician's arguments for the different opinion (similarly as the difference in opinions is resolved between two physicians) and then to recompute the estimates. It is however unclear how this can be accomplished in the current workflows and requires future research. These are important ethical issues of paramount importance to industry which need to be resolved with cooperation with stakeholders including clinical and ethical experts, patient advocacy groups, governmental bodies, and of course, the industry.[19] Finally, we anticipate that, when AI makes a clinical decision without human intervention some day perhaps in the not-too-distant future, we will face a complex problem of who is liable for a wrong decision made by AI, similar to self-driving car liability.

Failures

Critical failures

If algorithms do not perform according to specifications, it constitutes a major problem for vendors. For example, DL-based image reconstruction can be unstable resulting in severe artifacts.[20] This risk is often a consequence of the poor generalization of the AI algorithms and the fact that results presented by an AI algorithm are often not explainable. If a spectacular error is made by AI, it is very damaging to the perception of a product even if it works within specified characteristics. When publishing an article, the same penalty is applied if the algorithm had an error or a spectacular error. However, when we deploy an AI product, a spectacular failure could be much more detrimental to the trust in the algorithm. These types of errors, although very damaging, are very hard to mitigate with the current state of knowledge about neural networks. On a positive note, as much as critical failures of AI are damaging to its reputation, they are at least easily identifiable as errors. There are some safety features that can be used ("graceful failure"). For example, if we use AI to compute quantitative values and if the computed values are outside of the physiologic range, one may display a message that AI failed to compute the value rather than providing it to the user. For classification problems, these types of mitigations are much more difficult to implement. This is certainly important from an industry point of view and an important direction of future research.

Uncertainty estimation

Clear communication to the interpreting physician of uncertainty in the AI result is crucial in building trust in the AI system because, as previously discussed, no AI system will be perfect or able to handle the huge range of real-world inputs. It is not practical, or even possible, for AI developers to aim for a perfect result every time, so communication of uncertainty is of paramount importance.[21,22] Large uncertainty alerts about low confidence in provided inference. This is particularly true for nuclear imaging techniques which produce data with high noise compared with other modalities, and this noise may translate to uncertainty in reconstructed images and AI decisions. Suppose we develop an AI algorithm which automatically detects the volume of interest (VOI) of abnormal uptake of FDG, ideally the algorithm would also provide an estimation of uncertainty on the VOI size and SUV. This uncertainty can be expressed by providing a range of values that with a high likelihood contains the true value (confidence intervals). This can also be done using Bayesian approaches where each value of the volume or SUV is assigned probability of being true.[23] Estimation of such uncertainty can be accomplished with neural networks using approximations to Bayesian approaches[24,25] or some other approximate methods.[26,27]

Malicious AI, adversarial attacks

Another potential concern is that AI and DL methods either by accident or maliciously may introduce perturbations in the images. Some of these perturbations can be imperceptible to humans but may have a drastic effect on AI outcomes. For example, in an image manufactured by malicious AI, the analyzing AI may detect a tumor with 100% certainty which remains completely invisible to a human observer. For more on adversarial attacks refer to the study by Ma and colleagues.[28] When used for PET image reconstruction, AI may also introduce perturbations with image textures different from those obtained by standard iterative methods, which may be misinterpreted as abnormalities.

To Err is Human. How Does This Apply to AI?

Another issue that industry faces is to roll out products that will over time earn the trust of radiologists and nuclear medicine physicians and convince them to use algorithmic advice. We already drew the reader's attention to challenges associated with explainability. *Algorithm aversion* is another, potentially more serious, obstacle which may prevent seamless acceptance of AI solutions. Human decision-makers are averse to algorithmic predictions after seeing them perform; even with evidence of noninferiority of the AI algorithm, humans still tend to follow advice given by humans because people more quickly lose confidence in algorithms than in human forecasters after observing them repeating a mistake.[29] Algorithmic aversion may be a major obstacle to adoption of AI. AI algorithms used in augmenting human decision-making will likely have to be held to very high standards by enforcing interuser and extrauser reproducibility. If we can, we should also provide quantitative values with confidence measures (see also section Regulatory Pathways). Confidence is also important for yes/no or other classification decision tasks, and some type of confidence measures should always be provided.

LOOKING INTO THE FUTURE OF AI IN PET

Various academic medical centers worldwide are currently investing to incorporate AI in both research and clinical research settings as a prelude to AI-supported clinical workflows. For those applications, AI is largely used to scale and automate data analysis for large cohorts in multiyear studies whereby thousands of images are analyzed retrospectively. In clinical research, AI is typically used for clinical decision support as a "second opinion" to that of the clinician, to increase the saliency of structures and functions of interest in images while increasing efficiency of acquisition, and/or to identify possible regions or planes of interest in images so the clinician may improve diagnosis, increase efficiency, and minimize fatigue.[30–32]

The important question for the industry is how we bring AI into the clinical workflow in an efficient and scalable way. In section AI During PET Data Acquisition, we consider using AI during PET data acquisition. In section Vendor-Neutral Data-Processing Platforms, we explore new AI ecosystems already proposed elsewhere[33–35] and discuss how to leverage the uniqueness of PET raw data (eg, LM) in such ecosystems.

AI During PET Data Acquisition

AI offers a whole new array of promising approaches that have the potential to optimize the utility of PET imaging by adjusting controllable parameters based on the specific patient anatomy, patient physiology, and scanner type. The basic idea of how to achieve this is summarized in **Fig. 1**. We present the ability of AI algorithms to combine various types of information to provide just-in-time inferences which help to create high-fidelity PET data at the PET scanner while data are being acquired.

In this section, we provide example scenarios of how such AI inferences can be applied. In *scenario 1*, while the data are being acquired during a single-bed position, AI analyses the *data acquired* (**Fig. 1**) and uses criteria of *acceptable data quality* to determine if a sufficient number of counts were acquired up to this moment. An example of what problem this may partially solve is patient motion. If AI detects substantial patient motion, it triggers additional time for data acquisition also informing the operator. In *scenario 2*, suppose we scan a patient to determine whether the SUV in a given VOI changed versus the SUV measured in a previous PET scan. We provide the AI the *previous PET/CT scan*, *data acquired*, and maximum threshold for uncertainty of a decision (**Fig. 1**). We want to know if the SUV increased/decreased by 20%

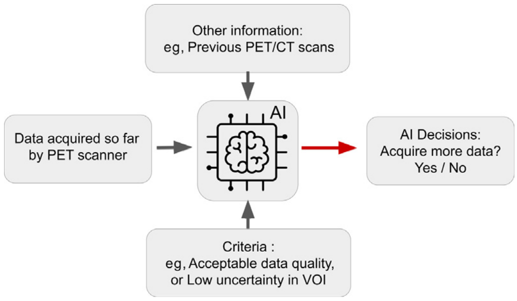

Fig. 1. Conceptual depiction of AI used at the PET scanner during data acquisition. *Gray arrows* indicate input to AI (data acquired so far, other data acquired in the past, and criteria for decision-making) and *red arrow* indicates output from AI.

with 95% certainty. AI analyzes the data and computes the maximum possible certainty that can be reached and the additional acquisition time to reach it.

The PET scanner is also a location where manufacturer-specific AI can be deployed. Once the raw data are created and the image is reconstructed, an AI algorithm can generate insights which can be sent to Picture archiving and communication system (PACS) or other destinations along with the data. Such solutions may be very effective as the manufacturer controls the type of data the AI algorithm is exposed to. The downside is of course that it is limited to individual manufacturers.

Vendor-Neutral Data-Processing Platforms

An effective approach to deployment of AI in radiology and other clinical environments is unclear. It is likely however that in the near future, we will have hundreds of AI algorithms approved for use in clinics and operating on different parts of clinical workflow and data. If we do not have a common platform to deploy them and rather depend on each AI vendor to use their own methods, the deployment and growth of AI in PET could stall as the complexity quickly becomes unmanageable.

To address this, new vendor-neutral data-processing platforms (VNDPs) are proposed.[33–35] In radiology, the VNDPs are interfaced with PACS. AI and other algorithms can process the data pulled from PACS and other hospital IT systems. After processing, the output can be sent back to the PACS, be saved on different archives, or displayed as shown in **Fig. 2**. We will not discuss these workflows in detail here and refer the reader to references available on this topic (vid.[33–35]). Software units that operate in VNDPs can be stacked together if their output/input type fits. Once stacked, blocks can be replaced by different blocks or stack of blocks. Such an architecture has a similarity with those used by smartphones[34] because software units are "sitting" on the platform and are activated if the "right" data arrives and they can be swapped/updated on user requests. Using this analogy, we will refer to the software units as "apps" (**Fig. 2**).

To provide an example of data processing in a VNDP platform, let us consider **Fig. 2** and processing by apps 4, 5, and 6. The input consists of PET/CT images. App 4 segments the liver using CT, app 5 detects liver lesions using PET and CT, and app 6 performs diagnosis and computes SUV using PET and CT if lesions were detected. Note that outputs from apps 4 and 5 are used by app 6.

Extended VNDP platform—processing standardized LM

Archiving PET data in a raw LM format has many advantages as it gives the ability to retrospectively reconstruct images on demand. There are many examples of where such flexibility is beneficial. For example, when training AI algorithms, it allows the developer to create a larger variety of images in terms of resolution and noise from just a single raw datafile. It also allows the developer to vary the total number of counts simulating different doses. The LM format may contain information about deposited energy and time-of-flight per event information, exact crystal pairs in which the gamma photons were detected, which may lead to development of improved reconstruction algorithms or correction algorithms compared with histogram (sinogram) data. As timing information is available for each event, it allows for various patient motion corrections.

The availability of LM data in new ecosystems would open opportunities to processing PET data, training new AI algorithms, and deriving AI inferences. The data reconstruction in such an ecosystem would just be another processing app which can be inserted in the processing pipeline (recon apps in **Fig. 3**). An example of such processing could be, for example, raw LM data

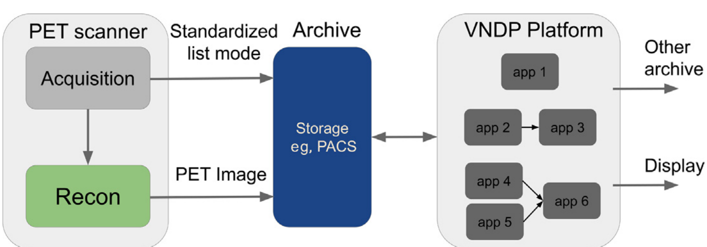

Fig. 2. Simplified PET data flows in new AI ecosystems. Data stored in, for example, PACS is pulled to the VNDP platform where they are processed by software units ('apps'). Apps can be used as a single processor (app 1) or stacked (eg, app 2 and 3 or apps 4, 5, and 6). VNDP platforms allow for creation of custom workflows with custom apps. Output from VNDP platform can be sent back to original storage, other archive, or displayed. Interactions with hospital information systems and other sources of information are omitted for clarity. Applications of AI before data reach storage are not shown.

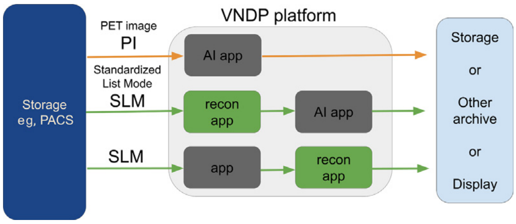

Fig. 3. PET data flow in VNDP. *Orange arrows* show dataflow in the new AI ecosystem with standard PET images processed by a single AI app. *Green arrows* show new PET-specific dataflows proposed in this work. SLM can be reconstructed by recon app and processed by AI app. In the third example, SLM is preprocessed (eg, randoms correction) and then reconstructed by the recon app. Interactions with hospital information systems and other sources of information are omitted for clarity.

correction for randoms or scatter which could be performed before image reconstruction. Ideally, in such an ecosystem, one would like to standardize the format of LM data to make it easier for vendors to develop apps which would work directly on LM data irrespective of the type of scanner the data were generated on. We refer to such a format as standardized LM (SLM) format. To perform state-of-the-art reconstruction, reconstruction applications need information including geometry, detector calibration, sensitivity, and so forth, which would have to be included in the SLM. We note that such a format does not exist at the time of writing this article as each scanner vendor uses a proprietary format. We note that standards for raw data are long established in Single Photon Computed Tomography (SPECT),[36,37] and more recently in MRI.[38] The SLM format for PET needs to be designed and approved by all stakeholders. A first step toward this goal would be that manufacturers disclose nonsensitive parts of their file formats, as some have already agreed to in the context of open-source projects.[39–41]

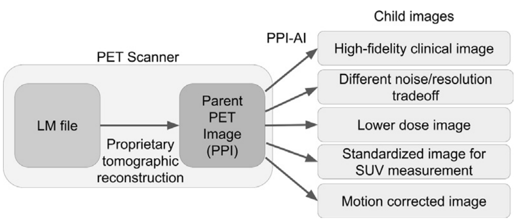

Fig. 4. LM data are reconstructed using vendor-specific proprietary software at the scanner. Each manufacturer creates manufacturer-specific AI models (PPI-AI) to transform the parent PET image to child images needed for various clinical and research tasks.

Extended VNDP platform—processing parent PET image

Although SLM in the VNDP platform provides enormous flexibility in constructing custom processing pipelines, handling LM files presents challenges. They are very large files (3–20 GB), and storage and network demands are considerable. Each vendor has a proprietary highly optimized software program which reconstructs images directly from the LM or sinograms created from the LM. Reconstruction software may have specific computing hardware requirements that may not be readily available in the VNDP platform.

AI offers an alternative approach toward creating a practical platform for generalization of the reconstruction process across different scanners and manufacturers without explicitly using LM files. The suggested approach gives up some generalizability compared with the SLM approach described in section 3.2.1, but it is more practical and well suited for use within a VNDP platform. We refer to this concept as the *parent PET image* (PPI) and summarize it in **Fig. 4**.

The main idea is that instead of handling SLM in the new ecosystem as shown in **Fig. 3**, we reconstruct on the scanner a parent image (or images) and use it instead of SLM in the VNDP platform. In the VNDP platform, PPI is then used to generate on demand various child images (**Fig. 4**). The generation of child images from the PPI is performed using deep CNNs referred to in this article as PPI-AI. The PPI-AI are types of apps in the ecosystem (**Fig. 2**) that convert PPIs to child images.

The PPI can be, for example, the high-fidelity image. PPI can actually also be a set of images, such as high-fidelity images with and without attenuation correction, resolution modeling, and so forth. If time-of-flight is available, it could also contain back-projections at different angles, as used by the DIRECT method.[42] There are many possibilities on how to define PPI, and research is needed to determine which of those choices would be optimal. The PPI is reconstructed on the scanner, and it is stored in PACS possibly along with some child images. The PPI can be pulled to the VNDP platform and almost instantaneously converted to any child image as the inference using the PPI-AI CNN model is fast. Once converted to a child image, it can be further processed by AI apps or other apps as a regular PET image (**Fig. 5**).

Looking at **Fig. 4**, the PPI can be converted to a high-fidelity image, the best utility image that a vendor can generate from the LM file. When training AI apps to be used in new ecosystems (**Fig. 2**), we would like to use images of various quality with various artifacts for the app to be more robust and general. A PPI-AI model can be

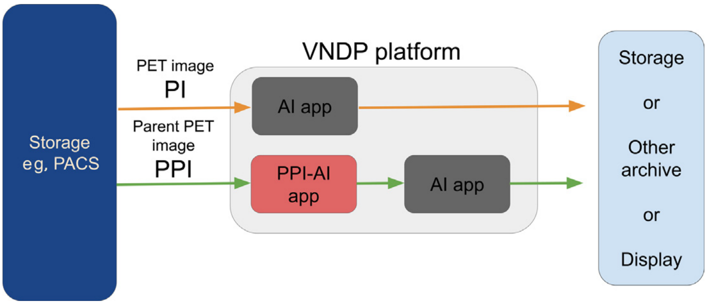

Fig. 5. PET data flow in VNDP with parent PET image (PPI) concept. *Orange arrows* show dataflow in the new AI ecosystem with standard PET images processed by a single AI app. *Green arrows* show new PET-specific dataflows proposed in this work. PET parent image is pulled from PACS and converted by PPI-AI app to a PET image which is processed by a single AI app. Interactions with hospital information systems and other sources of information are omitted for clarity.

trained to generate poorer quality images from the PPI. Examples of such are shown as different noise/resolution tradeoff and lower dose child images in **Fig. 4**.

There are ongoing efforts to harmonize and standardize results obtained on different scanners.[43,44] This can also be done using the PPI by creating harmonized child images. For this, we would require collaboration between vendors to create a single PPI-AI CNN model which could generate harmonized images from PPIs of different vendors. We can take this concept further and imagine a situation where the user points a cursor on a lesion when viewing a high-fidelity image, the standardized image is created in the background transparently to the user, and the viewing system displays standardized SUV values.

The fifth child image example provided in **Fig. 4** is physiologic motion (eg, respiratory) correction using PPI-AI. If no motion correction is applied, regions of the PPI with motion will appear blurry. PPI-AI models can be trained to recover resolution from blurred PPIs. Alternatively, a PPI could contain several images, for example, in different motion states, or one in end-expiration and one without motion correction, from which a fully motion-corrected image can be produced.

The training of PPI-AI models is conceptually straightforward. Suppose we want to create a PPI-AI model that generates from the PPI a half-dose image, first, we identify a training set which contains, for example, 1000 PET scans from some patient population. Then, we reconstruct those 1000 images from LM data using only half of the counts available in the LM. Then, we create PPIs by reconstructing images using all counts and high-fidelity reconstructions. We train neural networks (PPI-AI) with PPIs as the input and half-dose images as the target. This completes the creation of the PPI-AI model. Similarly, any other PPI-AI model can be trained. In the aforementioned steps, we assumed that high-fidelity reconstruction image is the PPI, but this may not necessarily be the optimal choice as already discussed.

A disadvantage of using the PPI compared with SLM is that the PPI contains less information than the LM file. Timing information is not available, and although the PPI can in general be a dynamic (or ECG-gated) sequence, it cannot be time reframed to a different sequence. We also do not have access to deposited energy, time of flight, and so forth. However, we remember that some of the information is transferred to PPI-AI models during training. Intuitively, during PPI-AI inference, when child images are generated from the PPI, not only the information in the PPI is used but also the information "stored" in PPI-AI models.

Another disadvantage of PPI is that if a manufacturer improves the tomographic reconstruction algorithm and wants to update it on the scanner, all PPI-AI models have to be retrained, which could be an automated process, but it is computationally intensive. If novel reconstruction is to be applied retrospectively to data acquired in the past, the PPIs have to be updated as well.

SUMMARY

In section Challenges for commercialization, we presented important challenges for creating and adopting AI solutions in clinics from the point of view of the industry. In section Looking into the future of AI in PET, we concentrated on PET explored unique to PET applications of AI during data acquisition. We examined a flexible and scalable ecosystem for deployment of AI and described a synergy of such systems with an idea of SLM data and the other solution presented here based on the PPI concept.

There are emerging new workflows and data ecosystems in radiology. In addition to facilitating AI deployment, they provide a tremendous opportunity for the PET community to transform the current paradigm of PET data processing.

CLINICS CARE POINTS

- Reconstruct PET data on demand (eg, just before or during reading) from raw data such as standardized list-mode data or parent PET images.
- Use raw data as a part of the "patient medical record."
- Include "DICOM push" for ease of raw data transfer/storage/management.
- Archive raw data which is essential for future improved reconstruction with motion correction, harmonization, or comparison with prior images.
- Process raw data in new AI ecosystems.

DISCLOSURE

S.A., E.A., A.C., A.I., and S.P. are employees of GE, Canon, United Imaging, NVIDIA, and Philips, respectively. K.T. is the owner of Algorithms and Software Consulting Ltd.

ACKNOWLEDGMENTS

Authors would like to thank Sven Zuehlsdorff, PhD, Siemens Medical Solutions USA, Inc., Hoffman Estates, IL, USA, sven.zuehlsdorff@siemens-healthineers.com for his input. The opinions expressed by authors in this article may not necessarily represent the official opinions of their employers. This publication is partly supported by the European Union's Horizon 2020 research and innovation programme under grant agreement Sano No 857533 and the International Research Agendas programme of the Foundation for Polish Science, co-financed by the European Union under the European Regional Development Fund.

REFERENCES

1. LeCun Y, Bengio Y, Hinton G. Deep learning. nature. 2015;521(7553):436–44.
2. Sharif MS, Abbod M, Amira A, et al. Artificial neural network-based system for PET volume segmentation. J Biomed Imaging 2010;4:2010.
3. Zhao X, Li L, Lu W, et al. Tumor co-segmentation in PET/CT using multi-modality fully convolutional neural network. Phys Med Biol 2018;64:015011.
4. Ypsilantis P-P, Siddique M, Sohn H-M, et al. Predicting response to neoadjuvant chemotherapy with PET imaging using convolutional neural networks. PLoS One 2015;10:e0137036.
5. Gong K, Kim K, Cui J, et al. The evolution of image reconstruction in PET: from filtered back-projection to artificial intelligence. PET Clinics. Current issue. doi:10.1016/j.cpet.2021.06.004.
6. Sanaat A, Zaidi H. Depth of interaction estimation in a preclinical PET scanner equipped with monolithic crystals coupled to SiPMs using a deep neural network. Appl Sci 2020;10(14):4753.
7. Berg E, Cherry SR. Using convolutional neural networks to estimate time-of-flight from PET detector waveforms. Phys Med Biol 2018;63(2):02LT01.
8. Gong K, Berg E, Cherry SR, et al. Machine learning in PET: from photon detection to quantitative image reconstruction. Proc IEEE 2019;108(1):51–68.
9. Sheller MJ, Edwards B, Reina GA, et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. Sci Rep 2020;10(1):1–2.
10. Sheller MJ, Reina GA, Edwards B, Martin J, Bakas S. Multi-institutional deep learning modeling without sharing patient data: A feasibility study on brain tumor segmentation. International MICCAI Brainlesion Workshop September 16-20, 2018 (pp. 92-104). 21st International Conference, Granada, Spain Springer, Cham.
11. Crimi A, Bakas S, Kuijf H, Menze B, Reyes M, editors. Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: Third International Workshop, BrainLes 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 14, 2017, Revised Selected Papers. Springer; 2018 Feb 16.
12. D'Amour A, Heller K, Moldovan D, et al. Underspecification presents challenges for credibility in modern machine learning. arXiv preprint arXiv:2011.03395. 2020 Nov 6.
13. European Commission. White Paper on Artificial Intelligence: A European Approach to Excellence and Trust. Report. 2020. Available at: https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf.
14. Artificial intelligence/machine learning (AI/ML)-Based. Software as a medical Device (SaMD) action plan. Available at: https://www.fda.gov/media/145022. Accessed July 22, 2021.
15. Proposed regulatory framework for Modifications to artificial intelligence/machine learning-based software as a medical Device. Available at: https://www.fda.gov/media/122535. Accessed July 22, 2021.
16. Arrieta AB, Díaz-Rodríguez N, Del Ser J, et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion 2020;58:82–115.
17. Wu N, Phang J, Park J, et al. Deep neural networks improve radiologists' performance in breast cancer screening. IEEE Trans Med Imaging 2019;39(4):1184–94.

18. Sitek A, Wolfe JM. Assessing cancer risk from mammograms: deep learning is superior to conventional risk models. Radiology 2019;292(1):67–8.
19. Fenech M, Strukelj N, Buston O. Ethical, social, and political challenges of artificial intelligence in health. London: Wellcome Trust Future Advocacy; 2018.
20. Antun V, Renna F, Poon C, et al. On instabilities of deep learning in image reconstruction and the potential costs of AI. PNAS 2020;117(48):30088.
21. Begoli E, Bhattacharya T, Kusnezov D. The need for uncertainty quantification in machine-assisted medical decision making. Nat Machine Intelligence 2019;1(1):20–3.
22. Kompa B, Snoek J, Beam AL. Second opinion needed: communicating uncertainty in medical machine learning. NPJ Digital Med 2021;4(1):1–6.
23. Sitek A. Data analysis in emission tomography using emission-count posteriors. Phys Med Biol 2012; 57(21):6779.
24. Neal RM. Bayesian learning for neural networks. New York: Springer Science & Business Media; 2012.
25. Wilson AG. The case for Bayesian deep learning. arXiv preprint arXiv:2001.10995. 2020 Jan 29.
26. Lakshminarayanan B, Pritzel A, Blundell C. Simple and scalable predictive uncertainty estimation using deep ensembles. arXiv preprint arXiv:1612.01474. 2016.
27. Abdar M, Pourpanah F, Hussain S, Rezazadegan D, Liu L, Ghavamzadeh M, Fieguth P, Khosravi A, Acharya UR, Makarenkov V, Nahavandi S. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. arXiv preprint arXiv:2011.06225. 2020 Nov 12.
28. Ma X, Niu Y, Gu L, et al. Understanding adversarial attacks on deep learning based medical image analysis systems. Pattern Recognition 2021;110: 107332.
29. Dietvorst BJ, Simmons JP, Massey C. Algorithm aversion: people erroneously avoid algorithms after seeing them err. J Exp Psychol Gen 2015;144(1): 114.
30. Demirer M, Candemir S, Bigelow MT, et al. A user interface for optimizing radiologist engagement in image data curation for artificial intelligence. Radiol Artif Intelligence 2019;1(6):e180095.
31. White RD, Erdal BS, Demirer M, et al. Artificial Intelligence to Assist in Exclusion of Coronary Atherosclerosis during CCTA Evaluation of Chest-Pain in the Emergency Department: Preparing an Application for Real-World Use. arXiv preprint arXiv:2008.04802. 2020 Aug 10.
32. Hashemian B, Manchanda A, Li M, et al. Review (2020). Clinical deployment and validation of a radiology artificial intelligence system for COVID-19.
33. Leiner T, Bennink E, Mol CP, et al. Bringing AI to the clinic: blueprint for a vendor-neutral AI deployment infrastructure. Insights Imaging 2021;12(1):1.
34. Enzmann DR, Arnold CW, Zaragoza E, et al. Radiology's information architecture could Migrate to one emulating that of smartphones. J Am Coll Radiol 2020;17(10):1299–306.
35. Allen B, Dreyer K. The artificial intelligence ecosystem for the radiological sciences: ideas to clinical practice. J Am Coll Radiol 2018;15(10): 1455–7.
36. National Electrical Manufacturers Association. NEMA PS3/ISO 12052, Digital Imaging and Communications in Medicine (DICOM) Standard. Available at: http://dicom.nema.org/medical/dicom/current/output/html/part01.html.
37. Todd-Pokropek A, Cradduck TD, Deconinck F. A file format for the exchange of nuclear medicine image data: a specification of Interfile version 3.3. Nucl Med Commun 1992;13(9):673–99.
38. Inati SJ, Naegele JD, Zwart NR, et al. ISMRM Raw data format: a proposed standard for MRI raw datasets. Magn Reson Med 2017;77(1):411–21.
39. Wadhwa P, Thielemans K, Efthimiou N, et al. PET image reconstruction using physical and mathematical modelling for time of flight PET-MR scanners in the STIR library. Methods 2021;185:110–9.
40. Ovtchinnikov E, Brown R, Kolbitsch C, et al. SIRF: synergistic image reconstruction framework. Computer Phys Commun 2020;249:107087.
41. Markiewicz PJ, Ehrhardt MJ, Erlandsson K, et al. NiftyPET: a high-throughput software platform for high quantitative accuracy and precision PET imaging and analysis. Neuroinformatics 2018;16(1):95–115.
42. Matej S, Surti S, Jayanthi S, et al. Efficient 3-D TOF PET reconstruction using view-grouped histo-images: DIRECT—Direct image reconstruction for TOF. IEEE Trans Med Imaging 2009;28(5):739–51.
43. Sullivan DC, Obuchowski NA, Kessler LG, et al. Metrology standards for quantitative imaging biomarkers. Radiology 2015;277(3):813–25.
44. Boellaard R, Delgado-Bolton R, Oyen WJ, et al. Fdg PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. Eur J Nucl Med Mol Imaging 2015;42(2):328–54.