Is there a role for image science in the brave new world of artificial intelligence?

Harrison H. Barrett

OPINION

Is there a role for image science in the brave new world of artificial intelligence?

Harrison H. Barrett[a,b,]*
[a]University of Arizona, Wyant College of Optical Sciences, Tucson, Arizona, United States
[b]University of Arizona, Center for Gamma-Ray Imaging, Department of Medical Imaging, Tucson, Arizona, United States

Abstract. Knowledge of the principles of image science is essential to the successful application of artificial intelligence in medical imaging. © 2019 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JMI.7.1.012702]

Keywords: image science; artificial intelligence.; published online Nov. 22, 2019.

1 Introduction

There is enormous current interest in the use of artificial intelligence (AI) in acquiring and interpreting medical images. As an example, the 2019 SPIE Medical Imaging conference was, as usual, divided into nine tracks representing different imaging technologies or clinical applications, but roughly half of all presentations, across all nine tracks, involved some form of AI. Evidently the current view is that AI is the solution to virtually any problem in imaging.

The same sentiment seems to run through almost all applications of AI. The details of data acquisition and preprocessing are dismissed as "domain knowledge," of no direct relevance to the inference. Given some number of exemplars of the data (often very few), it is tacitly assumed that a neural network or other AI-based inference engine can be trained to perform the inference task "better" in some sense than a more conventional method that does use domain knowledge. The justification of this assumption, if any is given, seldom seems convincing to this author.

The objective of this commentary is to survey some basic results from image science that may be useful in designing and validating AI-based imaging systems. Throughout, we use the term AI broadly, to include machine learning (ML), computer-aided diagnosis (CAD), deep learning (DL), generative adversarial networks (GANs), and other such methods.

For relevant background in imaging, see *Foundations of Image Science*[1] or any paper containing the acronym OAIQ (objective assessment of image quality).[2–6] For recent papers on AI in imaging, see the special issue of this journal on the topic[7] or recent Proceedings of the SPIE Medical Imaging Conference. For an enthusiastic review of deep learning in medical imaging, see Greenspan et al.[8] and the papers that follow it in another special issue. A comprehensive but readable book on deep learning is Goodfellow et al.[9]

2 Key Concepts from Image Science

2.1 Objects and Images

A fundamental error in image analysis is failing to distinguish objects and images. For example, one might segment an image with some algorithm and leap to the conclusion that the actual boundary of the object has been determined. In fact, the actual boundary may not even be defined unless the object is a geometric solid. For example, the interface between a tumor and surrounding normal tissue might be best described as a fractal surface, with an enclosed volume that depends on the measurement scale. More commonly, the density of cancer cells might fall off gradually rather than abruptly as one moves away from the center of a tumor; setting a threshold to define the boundary of the *image* of the tumor is again to ignore the distinction between object and image.

Similarly, one might calculate texture features of an image and infer erroneously that they reflect spatial inhomogeneities of the underlying object.[10] In tomographic imaging modalities, in particular, image texture can often be the result of noise in the raw projection data as modified by the reconstruction algorithm.[11,12]

Another common error in image-quality assessment is to assume, implicitly or explicitly, that a good image is one that accurately mimics the actual object, and hence one should compute the mean-square error (MSE) between an object and its image. There are many problems with this approach. The most fundamental one is that a digital image is a finite set of numbers on some grid, but the object being imaged is a function of continuous variables, such as $f(x, y, z)$. If the object function does not go to infinity anywhere and the object fits into a finite box (such as a CT scanner), then the function $f(x, y, z)$ can be treated as a vector $\mathbf{f}$ in a Hilbert space. In a mathematical sense, therefore, the object and image are vectors in different vector spaces, so their difference is not defined.

There are three ways to circumvent this problem: (a) sample the object function on the image grid; (b) interpolate the image so it better approximates the continuous object; or (c) ignore the problem and do a simulation where object and image are on the same grid.

The usual choice is (c), of course, but this still leaves another problem: How do we interpret the term "mean" in mean-square error? Again there are three options: (a) perform an average over pixels for one image of one (discrete) object; (b) perform an average over multiple realizations of a noisy image for a single object; or (c) average over both image noise and some ensemble of objects. Hence there are nine possible definitions of MSE; examples of all nine can be found in the literature. No justification for any of the choices is known to the author.

For AI applications to diagnostic medical imaging, the usual procedure is to acquire a labeled set of clinical images and use

them to train an AI system of some kind. In this approach, there is apparently no need to know what the true underlying object is. A problem with clinical images, however, is that the true diagnosis might be unknown or ambiguous, so simulated images are often used. In this case it is critical that the simulations be realistic, which means that both the object and the imaging system must be simulated accurately. This requirement, however, militates against the use of voxels (volume elements) in the object space. A human body is not a voxel array. The cardinal sin of image simulation is to use voxels for the object and make them comparable in size to the image pixels.

2.2 Tasks and Observers

The difficulties with MSE and other fidelity measures can be avoided with task-based assessment of image quality, a concept that should be familiar to readers of this journal. In brief, this approach to image-quality assessment requires specification of a task (the information desired from an image); the observer (how the task is performed), and the statistical properties of the object and image data (which limit the task performance). The observer in this paradigm can be a human, a computer program, a neural network or a mathematical construct called the ideal observer, which by definition achieves the best possible task performance for the specified task and statistical properties.

Much of the theory needed to perform task-based assessment of image quality was originally developed in the context of radar, where the task is implicit in the acronym: RAdio Detection And Ranging. In a simple radar system, a short pulse of microwave energy is generated and beamed into a medium that might contain targets of interest such as airplanes. If the target reflects sufficient pulse energy its return signal can be detected and the range to the target can be estimated from the round-trip transit time. The image data needed to perform the task are obtained by using multiple microwave pulses while the beam is scanned.

Receiver Operating Characteristic (ROC) curves were developed during World War II to quantify the detection performance of radar systems, and they were published in the open literature in the following decade.[13,14] Detection is performed by computing some functional of the radar data (the output of a matched filter, for example) and comparing it to a threshold. As most readers will know, an ROC curve is a plot of the probability of detection of a signal that is actually present versus the probability of a false alarm as the threshold is varied.

Given the ROC curve, there are many choices for a scalar figure of merit for detection performance. A common choice is area under the ROC curve, but another possibility is the probability of detection at a specific false-alarm rate. It is easy to see, however, that these pure ROC metrics are inadequate, even for the simple radar system described above. Radar requires range estimation as well as detection, and there may be a large and unknown number of targets in the field of view. In addition, modern radar systems can use Doppler methods to estimate the speed of each target, and they can use features of each return signal to distinguish different types of target.

Perhaps the earliest paper on ROC methods in medical imaging was by Lee Lusted in 1971.[15] Lusted considered human observers of medical images and only basic ROC detection methodology without additional estimation tasks. The meaning and use of area under an ROC curve was clarified by Hanley and McNeil,[16] and maximum-likelihood methods were devised by Dorfman and Alf[17] for estimating ROC metrics and their expected errors.

Around 1990 some of the deficiencies of simple ROC, notably the absence of any reference to estimation tasks, were beginning to be recognized, leading to many variants of the original ROC curve. In 2004, Chakraborty and Berbaum[18] summarized and compared a variety of methods of allowing an unknown number of targets per image and estimating the location of each. A major tool that emerged from these endeavors was the localization ROC (LROC) curve. Khurd and Gindi[19] developed the mathematics of LROC and derived an ideal observer for estimating the area under it. Eric Clarkson expanded this work to include more general estimation tasks.[20] He proposed an ROC-like curve, which he called EROC (estimation ROC), and he derived the corresponding ideal observer in that case. Subsequent work in many laboratories expanded the concepts of ROC and ideal observer much further and considered random backgrounds similar to textures observed in medical imaging.

ROC methods are now beginning to be applied to AI-based analysis of medical images, but it seems that they mainly consider simple detection-only task and original-recipe ROC. It is this author's opinion that the AI community would benefit greatly from reading the ROC literature of the last quarter-century and applying more modern ROC-like methodology to AI-based imaging systems.

2.3 Linear and Nonlinear Imaging Systems

From the considerations in Sec. 2.1, an image scientist (or any scientist who uses images) is well advised to treat an object as a function of continuous variables (which it is) and the resultant image as a discrete set of numbers (which it usually is, but see Sec. 2.9). In this viewpoint the imaging system is referred to as a continuous-to-discrete (CD) operator because it maps a function of continuous variables (equivalently, a Hilbert-space vector) to a discrete set of measurements.

The CD imaging operator, which we will denote as $\mathcal{O}$, can be either linear or nonlinear. By definition, a linear imaging operator must satisfy $\mathcal{O}[\alpha \mathbf{f}_1 + \beta \mathbf{f}_2] = \alpha \mathcal{O}[\mathbf{f}_1] + \beta \mathcal{O}[\mathbf{f}_2]$, where $\mathbf{f}_1$ and $\mathbf{f}_2$ are two different object vectors and α and β are scalar constants. In words, for a linear imaging operator, the image of a sum of two objects is the sum of the two images.

There are two forms of nonlinearity in imaging, which we can distinguish as intrinsic and extrinsic nonlinearities. Intrinsic nonlinearity is inherent in the imaging process. For example, x-rays are exponentially attenuated as they traverse tissue in CT, and in fluorescence lifetime imaging, the desired information—the lifetime—is a nonlinear functional of the fluorescence signal. In PET, SPECT and optical fluorescence imaging, on the other hand, the objects are concentrations of some radiotracer or fluorophore, which are directly observed in the imaging process, so intrinsic nonlinearity is not important.

Extrinsic nonlinearity arises from data processing after data acquisition. For example, modern tomographic imaging systems such as PET and SPECT often use iterative nonlinear algorithms to obtain an image. These algorithms are used to search for solutions that are consistent with the data in some sense and also satisfy certain physical constraints; for example, a concentration must be nonnegative.

For examples of the mathematical forms of both intrinsic and extrinsic nonlinearities, see Sec. 7.5 in *Foundations of Image Science*.[1] It is certainly clear that deep learning and other AI approaches can learn nonlinear functionals of training images,

but it does not appear that any work in this field has taken advantage of these known functional forms.

2.4 Null Functions and Nuisance Parameters

Two features of real imaging systems that are almost always ignored, in AI and other imaging applications, are null functions and nuisance parameters, but these features can be the main determinants of the performance of the imaging system, whether that performance is determined by AI or old-fashioned methods.

In brief, a null function (often called a ghost object) is a component of the object that makes no contribution to the image. A nuisance parameter, on the other hand, does contribute to the image data but not to performance of the task for which the image was acquired.

Almost all biomedical imaging systems exhibit null functions. One way to understand this point is to consider the input and output (i.e., the domain and range) of the operator $\mathcal{O}$. The input is a function in a Hilbert space, hence an infinite-dimensional vector. The output is an array of numbers, or an operator with rank no more than the number of measurements. Thus an infinite number of object functions can correspond to any particular data array, and there must be an infinite-dimensional null space.

If $\mathcal{O}$ is linear, then a method called singular-value decomposition (SVD), can be used to decompose any object into measurement and null functions: $\mathbf{f} = \mathbf{f}_{meas} + \mathbf{f}_{null}$ where $\mathcal{O}\mathbf{f} = \mathcal{O}\mathbf{f}_{meas}$ and $\mathcal{O}\mathbf{f}_{null} = 0$. See *Foundations of Image Science*,[1] Chapters 1 and 7, for details.

In some cases null functions are a form of aliasing. For many readers the term "aliasing" might conjure up the related concepts of bandlimited functions, Nyquist sampling and the Whitaker–Shannon sampling theorem, but for tomographic imaging more general approaches are needed.

The Whitaker–Shannon theorem requires that the functions being sampled must be bandlimited, but that means that the functions cannot also be spatially limited. To the contrary, objects considered in medical imaging must fit into the scanner, so they are spacelimited. This means that the 3D object can be represented exactly by a Fourier expansion, albeit with an apparently ridiculous number of terms. From this Fourier expansion we can form an even larger matrix, called the Fourier crosstalk matrix (FCM), which is an exact description of an arbitrary linear CD system with finite support. From the FCM we can compute expressions for task-based image quality, null functions, and practical reconstruction algorithms.

In tomographic systems, nuisance parameters arise most commonly from incorrect modeling of the system operator. In a collimator-based SPECT system, for example, penetration of gamma rays through the collimator septa might be a nuisance. Similarly, in PET or SPECT, failure to account for gamma rays scattered in the patient's body can lead to long tails on the reconstructed point-spread function.

The remedy for nuisance parameters in tomography is almost always better system modeling.

2.5 Estimation and Estimability

As noted in Sec. 2.2, the tasks for imaging systems are estimation, classification, or a combination of the two. In this section, we consider the role of null functions specifically for estimation tasks.

By way of example, consider the common task of estimating the amount of a tracer in some defined volume of a digital image, obtained, say, by PET or SPECT. At first blush this sounds trivial. One simply sums the reconstructed image voxel values over the selected region of interest (ROI), which often approximates a sphere, a cube or even a single voxel. When this method is attempted in practice, however, the accuracy of the estimation is often very poor. Many researchers explain that the errors are due to "partial-volume effects," which they then attempt to correct. The author of this opinion has no idea what a partial-volume effect is; it seems to have something to do with voxels in the object, but no such voxels exist.

The fundamental problem in this example is that the desired integral of the tracer distribution is not an estimable parameter. An estimable parameter is one for which there exists an unbiased estimate for all true values of the parameter. Bias is defined as the average (not mean-square) deviation of an estimate from the true value of the parameter over multiple trials. In imaging, a parameter of the object is a scalar-valued functional of the object, denoted $\Theta(\mathbf{f})$; this parameter is estimable for all $\mathbf{f}$ with a certain imaging system only if its value is independent of the null functions of the object for this system.

2.6 Noise and Task Performance

A fundamental misconception in much of the imaging literature is that noise in an image can be quantified by the variance at a point in the image. This statement would be valid for stationary, white, Gaussian noise, but the qualifiers rule out virtually all images. For tomographic images reconstructed with linear algorithms such as filtered back-projection, the noise may be approximately Gaussian by dint of the central-limit theorem, but it is far from stationary (independent of position in the image) or white (independent of spatial frequency). For nonlinear reconstruction algorithms such as MLEM (maximum-likelihood expectation-maximization), even the Gaussian property is lost because of the positivity constraint in the algorithm; log-normal statistics are common when images are reconstructed with the MLEM algorithm or its variants.

Correlations between random fluctuations at different points in an image play a critical role in determining task-based image quality in medical images for both human observers and machine observers. There are models of the human visual system that predict human task performance accurately, and there are many ways to compute the performance of the ideal observer. In these endeavors, the most common way of specifying the noise is in terms of covariance matrices.

There has been much less attention to trying to predict the task performance of AI systems for imaging applications. We note, however, that the AI performance depends not only on the task and noise properties of the data, but also on properties of the training data. There is a widespread intuition in the AI field that adding noise to the training data is equivalent to increasing the number of training samples. There is some anecdotal evidence of a small improvement in point variance by this so-called data-enrichment, but the author knows of no case where task performance has been improved.

A related point goes back to our discussion of objects and images in Sec. 2.1. If the objective of an AI system is deblurring, it would seem to be beneficial to train the system on high-resolution objects rather than blurred images. Indeed, a reasonable conjecture is that it is impossible for an AI system to recover fine details or high spatial frequencies that are not

represented in the training data. The author eagerly awaits a disproof of this conjecture.

2.7 Random Variables and Random Processes

We have emphasized that the objects of interest in real-world imaging are spatiotemporal functions, not arrays of voxels. Specifically in biomedical imaging, the functions are indicative of physiological or pathological processes in a patient. Though we hope to learn something about the patient by imaging methods, we can never hope to recover the functions in detail, so they must be described by the language of probability theory. We refer to the functions of medical or biological interest as physiological random processes (PRPs). Here the word "random" means simply "unpredictable"; a common synonym is "stochastic."

In elementary probability theory, the statistical properties of a continuous-valued scalar random variable are usually specified by a probability density function (PDF), but the same information can be conveyed by the characteristic function (ChFcn), which is the one-dimensional Fourier transform of the PDF.

The ChFcn formalism can also be applied to finite-dimensional random vectors such as images. If a random vector has M components (think detector pixels), then it is straightforward to define an M-dimensional PDF or the M-dimensional Fourier transform of the PDF, again referred to as the ChFcn.

Random processes are functions of continuous variables, hence vectors in an infinite-dimensional space. In this case it is very difficult to define a properly normalized PDF, but still straightforward to define its infinite-dimensional Fourier transform, which we now call a characteristic functional (ChFcnal).

Remarkably, there are large families of random processes for which the analytic form of the ChFcnal is known, generally with some freedom to tailor them to specific applications. Once this specialized ChFcnal is known, it is straightforward to compute finite-dimensional ChFcns for image data or for specific features of interest.

For more details on ChFcnals, see *Foundations of Image Science*[1] or the recent review by Clarkson and Barrett.[21] An extensive treatment of PRPs and ChFcnals in precision cancer therapy is given by Henscheid et al.[22]

2.8 Efficacy and Risk

In medicine, the use of ROC curves and their variants such as LROC and EROC is related to diagnostic tasks. There is also a burgeoning interest in ROC-like curves for therapy. The therapy operating characteristic (TOC) curve, a term first coined by Metz[23] in the context of radiation therapy, is a plot of probability of tumor control versus the probability of some critical adverse side effect as the radiation dose is varied. In the past decade, the TOC curve has been used for many radiation-therapy applications with both external and internal radiation sources, and it has been extended to chemotherapy.[24]

These ROC-like curves are all plots of the probability of a favorable outcome (e.g., tumor detection, tumor control) versus the probability of an unfavorable outcome (false alarm, damage to a normal organ) as something (detection threshold, radiation dose) is varied. In each case the vertical axis can be called an efficacy, and the horizontal axis is a risk to the patient (risk of missing a tumor or damaging a normal tissue).

Additional kinds of ROC-like curves can be generated by considering more general forms of efficacy and new forms of patient risk. The roadmap for this kind of investigation was developed by Fryback and Thornbury,[25,26] who identified six stages of efficacy:

1. technical capacity,
2. diagnostic accuracy,
3. diagnostic impact,
4. therapeutic impact,
5. patient outcomes, and
6. societal outcomes.

For more details on these forms of efficacy, see the extensive review article on task-based measures of image quality and their relation to radiation dose and patient risk.[27] This reference also shows how different levels of efficacy can lead to new image-quality metrics and new kinds or ROC-like curves.

All of these approaches to efficacy-based assessment of image quality can also be applied to AI systems. At the least, considering new forms of efficacy and corresponding new tasks might help to move the AI world away the search for a universal AI inference engine and into the direction of solving meaningful medical problems with realistic data.

2.9 Photon Counting and Photon Processing

The picture painted so far in Sec. 2 is that task-based assessment of image quality might ultimately be limited by null functions. We argued that real objects of biomedical interest are vectors in an infinite-dimensional space, and real digital images are finite sets of numbers, hence vectors in a finite-dimensional space. The difference in dimensionality implies, we asserted, that the imaging system must have a null space. In this subsection we explore a possible exception to this assertion.

The historical thread to be traced here begins with the classic Anger scintillation camera. developed over 60 years ago for imaging gamma rays in nuclear medicine – and still in wide use today. The earliest Anger cameras used a single-crystal sodium iodide (NaI) scintillator and a hexagonal array of seven photomultiplier tubes (PMTs). Each gamma-ray photon that was absorbed in the NaI produced a flash of light that spread out and produced signals in all seven PMTs. A capacitor array then functioned as an analog computer to estimate the x-y coordinates of the scintillation flash. The device was thus a photon-counting gamma-ray camera with analog estimation of x-y position of each event.

Inevitably, the analog computer was replaced by a digital one, the NaI crystal became much larger and the number of PMTs increased accordingly. Most importantly for this discussion, users saw opportunities to get more and better information from each scintillation event. New search algorithms and accurate camera calibration methods allowed researchers at the University of Arizona to get high-precision maximum-likelihood estimates (MLE) of gamma-ray energy and depth of interaction of each gamma-ray photon, as well as improved estimates of the x-y position of each event.[28] To avoid losing this precision in the data storage, the digital estimates were stored an a list, where each entry consisted of the four high-precision estimates (x-y-z-E) for each detected gamma-ray photon, without any binning. These four estimates are referred to as the attributes of each photon absorption event. For single-photon imaging (SPECT) we have four attributes per event, but for coincidence

imaging in PET with two detectors in play we can add time of arrival for each gamma ray and get as many as 10 attributes per positron annihilation.

This storage mode, referred to as list mode, turned out to have some important practical and theoretical advantages.[6,29] There was no loss of information, in the sense of reduced detector performance, if MLE and listmode storage were used. We showed in a recent publication[27] that MLE/listmode had to be a component of the ideal dose utilizer, which obtains best performance for any task in imaging with ionizing radiation. We refer to image detectors that satisfy these conditions as photon-processing detectors rather than photon-counting ones.

In simulations of tomographic imaging with alpha or beta particles, where we can also estimate direction of travel of each particle, we have demonstrated that the null functions nearly disappear if we have 4-6 attributes per event.[30,31]

3 Computational Methods for Image Science

To the unitiated, concepts from image science may seem arcane, even bewildering. How do we actually compute the performance of an ideal observer or the null functions of a particular imaging system? How do we handle tasks that require detection of a complex signal followed by estimation of some signal parameters? For any task, how do we assess the statistical significance of our results?

A good place to start reading about these issues is Barrett and Myers, *Foundations of Image Science*,[1] especially Chaps. 13 and 14. Chapter 13 covers the basics of statistical decision theory, with many worked examples, and Chapter 14 applies these methods to image quality. Later chapters work through the application of the methods in 13 and 14 to specific imaging systems, again with worked examples.

Foundations was finalized in 2003, and there have been many new developmens in computational image science since then. For example, Foundations placed considerable emphasis on continuous-to-discrete models of digital imaging systems, which seemed in 2003 to be the only possible approach. Now the new field of photon-processing detectors is leading to continuous-to-continuous (CC) systems with advantages not only in null functions as in Sec. 2.9, but also in sensitivity and spatial resolution. For some recent results, see Caucci.[32]

A long review article on radiation dose and risk for imaging with ionizing radiation[27] provides a general review of methods of computing image-quality metrics as a function of radiation dose. This paper also develops several new graphical methods for depicting the tradeoffs among radiation dose, image quality and patient risk.

Eric Clarkson and coworkers have developed many novel uses for Fisher information matrices (FIMs) in image-quality assessment. The FIM is traditonally used in estimation problems, but Clarkson showed that it is also useful for approximating the ideal observer for classification problems.[33,34] Another powerful tool for computing the ideal observer and its task-based performance is Markov-chain Monte Carlo simulation.[35] Finally, a pervasive problem in image-quality assessment is that inverses of very large matrices may be needed, leading to a malady called megalopinakophobia (fear of large matrices); cures for this disease are discussed in Foundations.[1]

Software for task-based assessment of image quality with both human and model observers has been developed at FDA; it can be found on Github at DIDSR/IQmodelo.

4 Maxims and Minims

4.1 *Maxims*

Here we present some maxims that one should observe in designing, evaluating and using any medical imaging system, whether or not it uses AI methodology in any sense.

- Many simulated imaging systems have no null functions; this means they have no relevance to real systems.
- If you must use MSE, be sure to specify which of the nine possible definitions you choose. And why.
- In tomography, don't forget the reconstruction algorithm; it may control the properties of the image more than the object does.
- Imaging systems include image detectors; models of imaging systems must do the same.
- Study nothing; i.e., characterize the null space of your imaging system.
- Decide what nuisances (parameters) arise in real life, and how you will deal with them.
- Enumerate all sources of randomness in your problem. Which dominate?
- Decide on a clinically relevant task.
- Choose the proper operating characteristic for your application. Turn it into a scalar figure of merit.
- If you want to solve an inverse problem, concentrate on the forward problem. Take advantage of everything about your data that is known from physics and mathematics.
- Embrace OAIQ. Do not substitute fidelity measures or point statistics for task performance. Why was the image acquired in the first place?
- OAIQ is inherently stochastic. One image tells you nothing at all. Especially if it has an arrow pointing to a tumor.

4.2 *Minims*

The term minim is used in many contexts to mean something very small. In music it is a half note. As a liquid measure a minim is 1/60th of a dram, which of course means it is 1/20th of a scruple. In the world of British apothecaries, it serves as a standard drop. In this paper it is a suggested minimal standard for designing and evaluating AI-based imaging systems.

- Eschew toy problems. Don't pretend that some collection of geometric shapes is a surrogate for a medical image.
- Do a background check. Fine anatomical structure in an image can be treated as a random background; it should not be ignored.
- Seek realism. If you are simulating clinical images for training or testing, do so realistically.
- Don't use apples to classify oranges. If clinical training data come from different sources, pay attention to the properties of the imaging hardware and algorithm at each.
- Be very skeptical of correlations. Is there any physical mechanism by which expression of a certain gene can affect a CT scan? The null hypothesis should be that it cannot.

- Explore conventional alternatives to machine learning. If you want to "learn" the Radon transform, for example, why not read the literature? Hint: it starts in 1917.

- Seek the ideal. Compute the performance of the ideal observer, or at least a lower bound to it. A good AI system might beat that lower bound, but if your AI system beats the IO itself, you are doing something wrong.

5 Conclusions

To answer the question in the title, we might pose the converse: Can there be a scientific approach to AI in imaging without image science? To paraphrase Descartes, "Dubito, ergo sum" (I doubt, therefore I am).

Readers who do not yet share these doubts are invited to study the two bullet lists above and see which, if any, of the issues highlighted there can be addressed satisfactorily without appeal to the theory of image science. Other readers may contribute to developing a scientific approach to AI in imaging by finding additional items to add to one or both lists.

Disclosures

The author has no competing interests to declare.

References

1. H. H. Barrett and K. J. Myers, *Foundations of Image Science*, John Wiley and Sons, Hoboken, New Jersey (2004).
2. H. H. Barrett, "Objective assessment of image quality: effects of quantum noise and object variability," *J. Opt. Soc. Am. A* **7**(7), 1266–1278 (1990).
3. H. H. Barrett et al., "Objective assessment of image quality II: fisher information, Fourier crosstalk, and figures of merit for task performance," *J. Opt. Soc. Am. A* **12**(5), 834–852 (1995).
4. H. H. Barrett, C. K. Abbey, and E. Clarkson, "Objective assessment of image quality III: ROC metrics, ideal observers and likelihood-generating functions," *J. Opt. Soc. Am. A* **15**(6), 1520–1535 (1998).
5. H. H. Barrett et al., "Objective assessment of image quality IV: application to adaptive optics," *J. Opt. Soc. Am. A* **23**(12), 3080–3105 (2006).
6. L. Caucci and H. H. Barrett, "Objective assessment of image quality V: photon-counting detectors and list-mode data," *J. Opt. Soc. Am. A* **29**(6), 1003–1016 (2012).
7. E. A. Krupinski, P. E. Kinahan, and P. La Riviere, "Special section guest editorial: artificial intelligence in medical imaging," *J. Med. Imaging* **6**(1), 011001 (2018).
8. H. Greenspan, B. Van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique," *IEEE Trans. Med. Imaging* **35**(5), 1153–1159 (2016).
9. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, Cambridge, Massachusetts (2016).
10. F. J. Brooks, "On some misconceptions about tumor heterogeneity quantification," *Eur. J. Nucl. Med. Mol. Imaging* **40**(8), 1292–1294 (2013).
11. H. H. Barrett, D. W. Wilson, and B. M. Tsui, "Noise properties of the EM algorithm. I. Theory," *Phys. Med. Biol.* **39**(5), 833–846 (1994).
12. D. W. Wilson, B. M. Tsui, and H. H. Barrett, "Noise properties of the EM algorithm. II. Monte Carlo simulations," *Phys. Med. Biol.* **39**(5), 847–871 (1994).
13. P. Woodward, *Probability and Information Theory: With Applications to Radar*, Number 4, Electronics and Waves Series, McGraw-Hill, New York (1953).
14. W. Peterson, T. Birdsall, and W. Fox, "The theory of signal detectability," *Trans. IRE Prof. Group Inf. Theory* **4**(4), 171–212 (1954).
15. L. B. Lusted, "Signal detectability and medical decision-making," *Science* **171**(3977), 1217–1219 (1971).
16. J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology* **143**(1), 29–36 (1982).
17. D. D. Dorfman and E. Alf Jr., "Maximum-likelihood estimation of parameters of signal-detection theory and determination of confidence intervals–rating-method data," *J. Math. Psychol.* **6**(3), 487–496 (1969).
18. D. P. Chakraborty and K. S. Berbaum, "Observer studies involving detection and localization: modeling, analysis, and validation," *Med. Phys.* **31**, 2313–2330 (2004).
19. P. Khurd and G. Gindi, "Decision strategies maximizing the area under the LROC curve," *IEEE Trans. Med. Imaging* **24**(12), 1626–1636 (2005).
20. E. Clarkson, "Estimation receiver operating characteristic curve and ideal observers for combined detection/estimation tasks," *J. Opt. Soc. Am. A* **24**(12), B91–B98 (2007).
21. E. Clarkson and H. H. Barrett, "Characteristic functionals in imaging and image-quality assessment: tutorial," *J. Opt. Soc. Am. A* **33**(8), 1464–1475 (2016).
22. N. Henscheid et al., "Physiological random processes in precision cancer therapy," *PLoS ONE* **13**(6), e0199823 (2018).
23. C. I. Metz et al., "Maximum likelihood estimation of dose-response parameters for therapeutic operating characteristic (TOC) analysis of carcinoma of the nasopharynx," *Int. J. Radiat. Oncol. Biol. Phys.* **8**(7), 1185–1192 (1982).
24. H. H. Barrett et al., "Therapy operating characteristic (TOC) curves and their application to the evaluation of segmentation algorithms," *Proc. SPIE* **7627**, 76270Z (2010).
25. D. G. Fryback and J. R. Thornbury, "The efficacy of diagnostic imaging," *Med. Decis. Making* **11**(2), 88–94 (1991).
26. J. Thornbury, "Eugene W. Caldwell Lecture. Clinical efficacy of diagnostic imaging: love it or leave it," *Am. J. Roentgenol.* **162**(1), 1–8 (1994).
27. H. H. Barrett et al., "Task-based measures of image quality and their relation to radiation dose and patient risk," *Phys. Med. Biol.* **60**(2), R1–R75 (2015).
28. H. H. Barrett et al., "Maximum-likelihood methods for processing signals from gamma-ray detectors," *IEEE Trans. Nucl. Sci.* **56**(3), 725–735 (2009).
29. H. H. Barrett, T. White, and L. C. Parra, "List-mode likelihood," *J. Opt. Soc. Am. A* **14**(11), 2914–2923 (1997).
30. Y. Ding, L. Caucci, and H. H. Barrett, "Charged-particle emission tomography," *Med. Phys.* **44**(6), 2478–2489 (2017).
31. Y. Ding, L. Caucci, and H. H. Barrett, "Null functions in three-dimensional imaging of alpha and beta particles," *Sci. Rep.* **7**(1), 15807 (2017).
32. L. Caucci et al., "Towards continuous-to-continuous 3D imaging in the real world," *Phys. Med. Biol.* **64**, 185007 (2019).
33. E. Clarkson and F. Shen, "Fisher information and surrogate figures of merit for the task-based assessment of image quality," *J. Opt. Soc. Am. A* **27**(10), 2313–2326 (2010).
34. S. Park and E. Clarkson, "Efficient estimation of ideal-observer performance in classification tasks involving high-dimensional complex backgrounds," *J. Opt. Soc. Am. A* **26**(11), B59–B71 (2009).
35. M. A. Kupinski et al., "Ideal-observer computation in medical imaging with use of Markov-chain Monte Carlo techniques," *J. Opt. Soc. Am. A* **20**(3), 430–438 (2003).

Harrison H. Barrett, PhD, is Regents Professor Emeritus of optical sciences and medical imaging at the University of Arizona. He is a fellow of OSA, APS, AIMBE, SPIE, and IEEE, and he is a member of both the National Academy of Engineering and the National Academy of Inventors. *Foundations of Image Science*, a 2004 book coauthored with Kyle J. Myers, was awarded the first SPIE/OSA J. W. Goodman Book Writing Award. Barrett has also received the IEEE Medal for Innovations in Healthcare Technology, the SPIE Gold Medal of the Society, the SNMMI Abersold Medal, OSA Mees Medal, and an honorary doctorate from University of Ghent. Most recently he was co-recipient of the inaugural SPIE Harrison H. Barrett Award in Medical Imaging (with A. E. Burgess, C. E. Metz, and R. F. Wagner).