Advanced intelligent Clear-IQ Engine (AiCE) Interpretable Model with Robust and Generalized Performance: Beyond Brain and Knee

Hung P. Do, PhD
Manager Medical Affairs – Clinical Scientist
Canon Medical Systems USA

Dawn Berkeley
Manager Medical Affairs – Clinical Development
Canon Medical Systems USA

Interpretable Model, Explainable Performance, and Rigorous Validations

Deep Learning[1] has been applied in many stages in radiology including administrative assistance, automatic and assisted scanning, data acquisition and image reconstruction, automatic detection, classification, quantification, image interpretation, diagnosis and prognosis, and clinical decision support. Although disease detection, quantification, and interpretation are popular AI-based solutions[2,3], the data acquisition and image reconstruction play an important role in the whole radiology workflow since accurate and high quality reconstructed images would provide confidence and improved performance of the downstream tasks.

Canon Medical Systems has introduced Advanced intelligent Clear-IQ Engine (AiCE) deep learning reconstruction (DLR)[4–6], a novel image reconstruction technique based on deep convolutional neural network (CNN), which is trained to differentiate Gaussian noise from signal and effectively removes noise while maintaining anatomical and pathological integrity allowing the reader to "see through the noise." AiCE performance has been demonstrated to be robust in clinical practice and generalized to variations in noise level, contrast, sequence parameter, scan protocol, anatomy, body habitat, and field strength.

Deep learning has successfully evolved to its current flourishing state due to three main components, which are (i) the availability of large high quality and clean data, (ii) advances in algorithm development, and (iii) the availability of cheap high-powered parallel computing. Among the three components, high-powered parallel computing has become ubiquitous and accessible thanks to the advent of graphics processing unit (GPU). The first two components were leveraged in the development of AiCE. First, the AiCE training data was not just standard quality images but they were carefully and intensively prepared to achieve exceptional image quality. Specifically they were acquired with 10 averages resulting in very high signal-to-noise ratio (SNR) and high image quality that was not clinically practical. Second, the AiCE network architecture was designed based on a combination of domain expertise (in medical imaging, magnetic resonance physics, and signal processing) and innovative deep learning algorithm allowing the AiCE model to be interpretable. The AiCE model interpretability and rigorous validations allow its robust and generalized performance to be explainable. Validations include bench testing, model observer study, and human observer study.

AiCE Model Interpretability

Deep learning model is referred as a black-box model since it is often seen as a process in which a computational engine (with a huge number of parameters) transforms inputs to desired outputs. The transformation and the computation processes are too complex and opaque to comprehend hence the name "black-box" model. AiCE model, on the other hand, was designed with the basis of magnetic resonance physics, medical imaging, and signal processing, as shown in Figure 1, where its components can be interpreted based on prior knowledge of physics and signal processing.

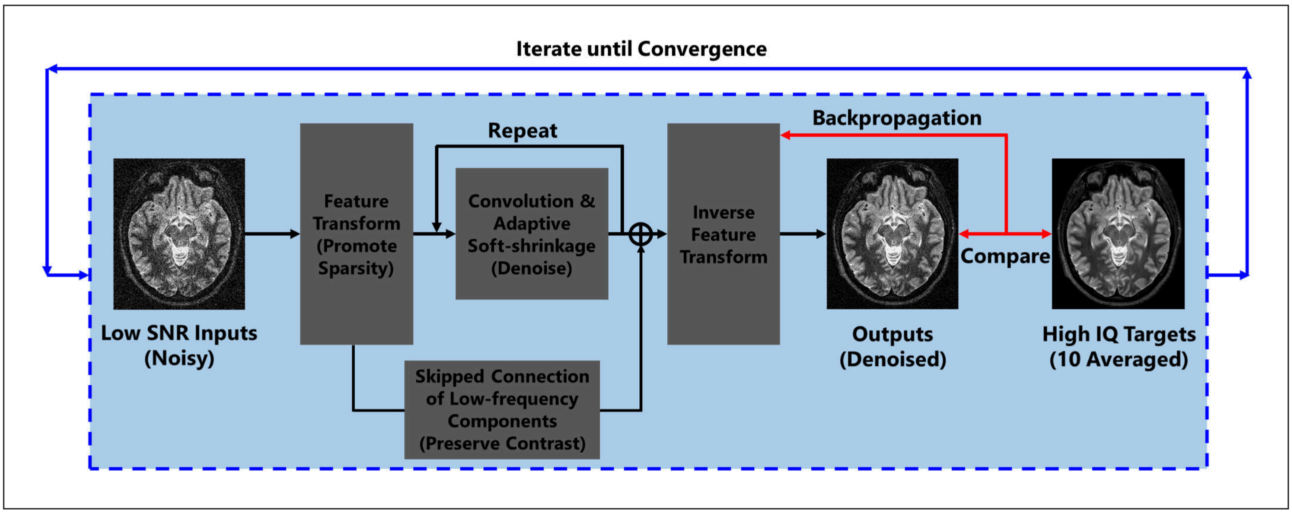

Figure 1 AiCE's Interpretable model

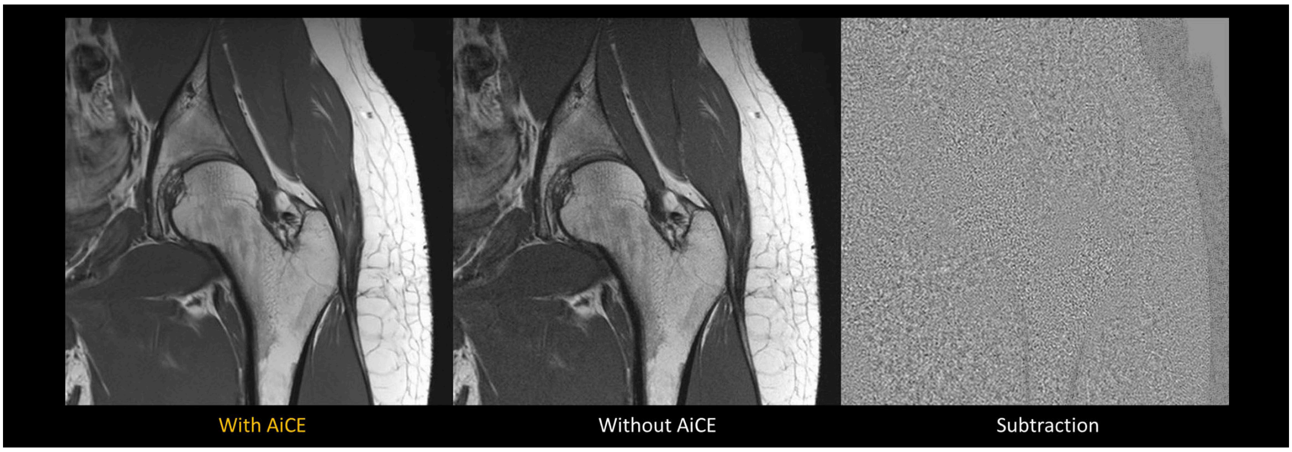

Figure 2 Subtracted image contains only noise.

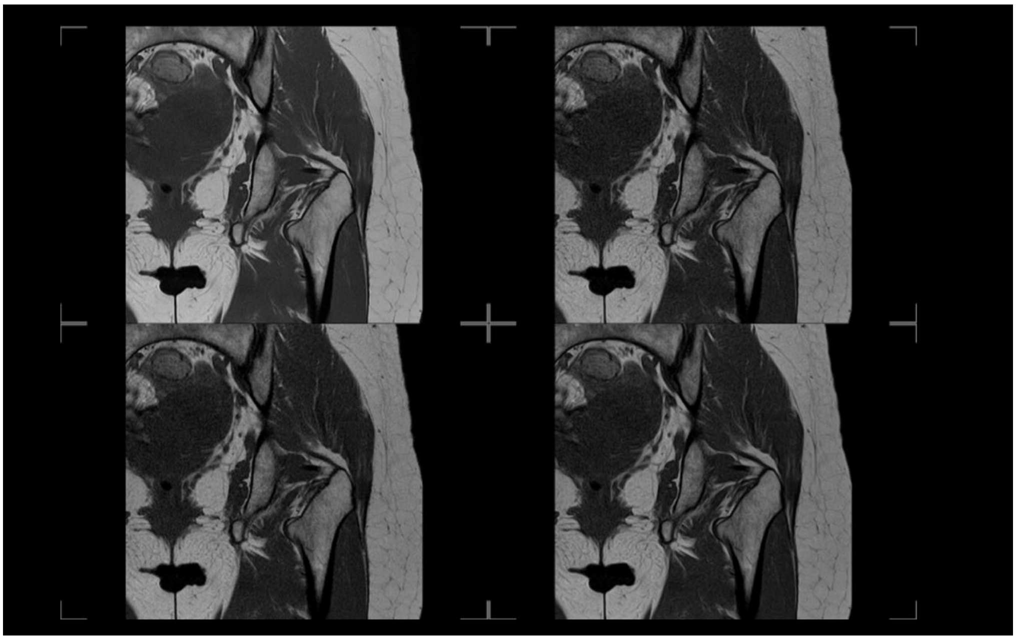

Figure 3 For each sequence, four randomized label-removed reconstructions (AiCE, NL2, GA43, and GA53) were prepared and shared with radiologists for review.

Bench Testing

The essence of training a deep learning model is not to best fit the training data but to generalize (i.e. to perform well) to the real-world data (i.e. the data that is unseen during model training and tuning). The phenomenon in which the deep learning model performs well on the training data and validation data but does poorly on the unseen data is called overfitting. During model development, a learning curve analysis is used to identify overfitting (i.e. poor generalization). The learning curve for the test data (i.e. the data that is not seen by the model during training and tuning) is calculated and visualized to confirm that the learning curve of the test data is monotonically decreased.

Common quantitative metrics such as root mean square error (RMSE), peak signal-to-noise ratio (PSNR), and structure similarity index measure (SSIM) are often used in traditional MR image reconstruction and also in machine learning and deep learning based MR image reconstruction literature to evaluate model performance and quality of reconstructed images. These metrics, however, are global and fail to capture local image quality and fidelity, which is important to medical imaging applications. For AiCE, image subtraction (i.e., an image without AiCE is subtracted from an image with AiCE) can be calculated to evaluate how well AiCE performs both globally and locally. AiCE was designed to only remove noise from the input image therefore the subtracted image is expected to only contain noise, ensuring that the anatomical and pathological features are maintained. An example of subtracted image is shown in Figure 2, where the subtracted image only contains noise. Additionally, model observer and human observer studies were performed to further evaluate model stability, robustness, and generalization beyond global quantitative metrics and bench testing.

Model Observer Study

Model observer (MO)[7,8] is an important tool for evaluating a reconstruction method in medical imaging since it is shown that MO performance is correlated with human observer performance.[9,10] A rigorous MO study often involves (i) identifying a clinically relevant task such as classification, estimation, or detection, (ii) specifying the population, (iii) selecting an observer, which is an appropriate mathematical model for the chosen task, and (iv) defining a figure of merit, which tells how well the observer performs. MO study could evaluate the reconstruction method's performance in terms of robustness and stability due to variations in inputs such as noise level, contrast level and size of low contrast objects, and also variations of different settings of the

	Overall Image Quality Assessment				Assessment of Specific Anatomical & Pathological Features			
Score	**Overall Image Noise**	**Overall Image Sharpness**	**Overall Image Degradation/ Artifacts**	**Overall Diagnostic Confidence**	**Features Sharpness**	**Features Contrast**	**Features Conspicuity**	**Features Diagnostic Confidence**
1	All structures are too noisy	All structures are not sharp on most images	All structures are degraded	Diagnostic confidence is very poor	Sharpness of features is very poor	Contrast between features and surrounding is very poor	Features conspicuity is very poor	Diagnostic confidence for the reviewed features is very poor
2	Most structures are too noisy	Most structures are not sharp on some images	Most structures are degraded	Poor	Poor	Poor	Poor	Poor
3	A few structures are noisy on most images	Most structures are sharp on some images	A few structures are degraded on most images	Fair	Fair	Fair	Fair	Fair
4	A few structures are noisy on a few images	Most structures are sharp on most images	A few structures are degraded on a few images	Good	Good	Good	Good	Good
5	There is no appreciable noise on any of the relevant images	All structures are sharp on all relevant images	There is no appreciable image degradation on any of the relevant images	Excellent	Excellent	Excellent	Excellent	Excellent

Table 1 Scoring criteria and instructions.[12] In addition to overall image quality assessment, specific anatomical and pathological features were also evaluated and scored by radiologists. Clinically relevant anatomical features were identified beforehand while the pathological features and findings were identified by radiologists during the review process.

reconstruction technique. For example, tens of thousands of instances (i.e. reconstructed images with various combinations of conditions and parameters) can be easily evaluated by the MO but it is impractical for human observer. It is even more important to perform such stability testing since deep learning based reconstruction method is shown to be susceptible to subtle variations in the input during deployment.[11] Additionally, a MO study could be used to evaluate a proposed reconstruction method against established predicate methods for specific clinically relevant task such as low contrast detection.

MO studies were carried out to compare the low contrast detection performance of AiCE and NL2 (i.e. a predicate method). The results show that AiCE performs similarly or better compared to the predicate method for the low contrast detection task.

Human Observer Study

An important step in ensuring efficacy and safety of a reconstruction method is to perform randomized blinded human observer study in which clinical data were prospectively collected. The data were collected using clinically relevant protocols from multiple scanners, field strengths, anatomies, and geographic locations. For each sequence, AiCE and three typical predicate methods (NL2, GA43, and GA53) were performed resulting 4 different reconstructions. The labels of these 4 reconstructions were removed and they were randomly ordered and shared with ABR-certified radiologists to review and score. An example of 4 reconstructions on a sequence in the hip protocol is shown in Figure 3.

Clinical Utilizations

AiCE was trained to intelligently and adaptively remove noise while maintaining structure integrity allowing one to "*see through the noise*." The SNR gained could be used for acquisition with improved resolution and/or shortened scan time. Furthermore, AiCE may enable acquisition of high-field liked image quality using the lower field scanner without the challenges associated with imaging at higher field such as (i) higher equipment and operating cost, (ii) higher specific heat absorption rate (SAR), and (iii) more artifacts due to higher B0 and B1 field inhomogeneities. Figures 6-13 are representative images from various anatomies that demonstrate the robustness, generalization, and potential benefits of AiCE in clinical setting. Figures 14-16 show comparison between 3T images and AiCE-powered 1.5T images acquired with the same resolution and similar scan time while other parameters were optimized for appropriate contrast at different field strength.

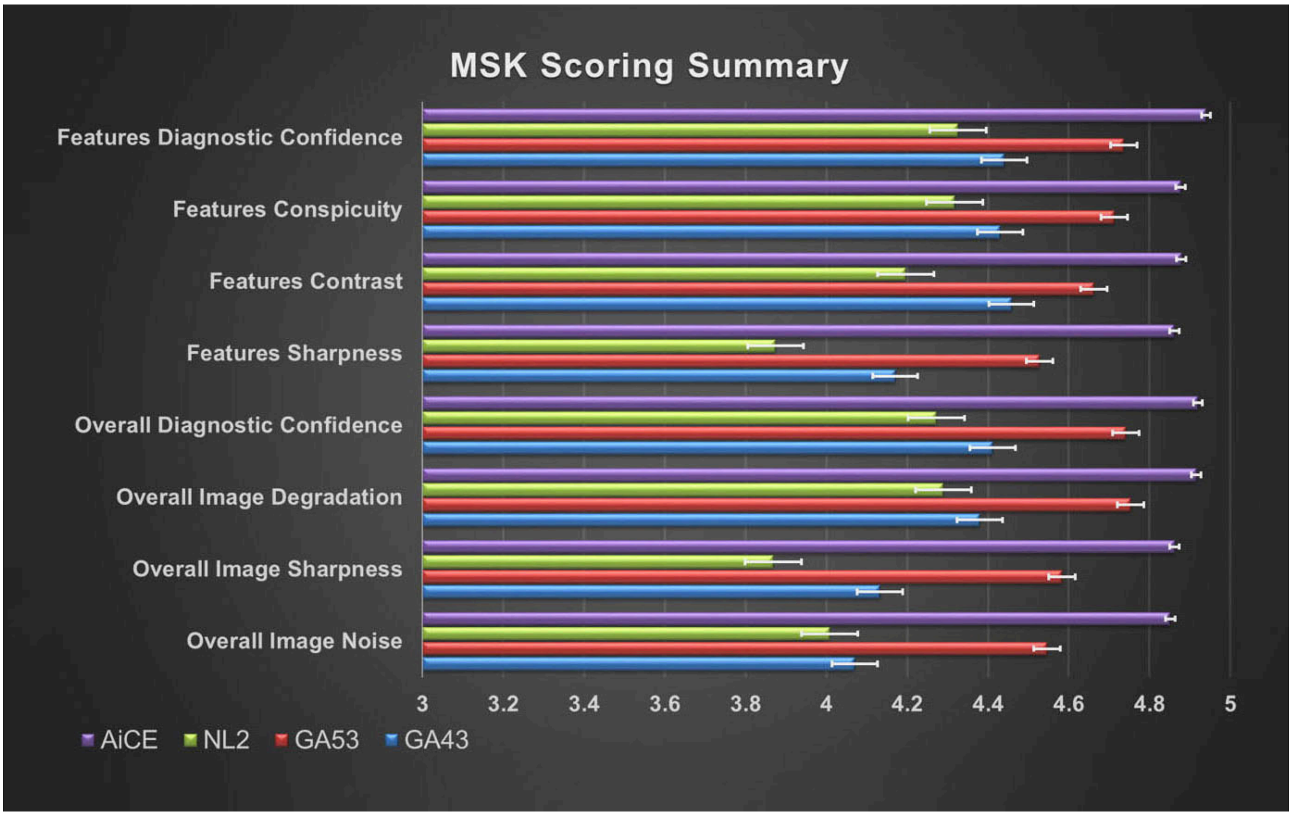

Figure 4 Image quality scores for MSK at 1.5T. AiCE has the highest scores in all categories.

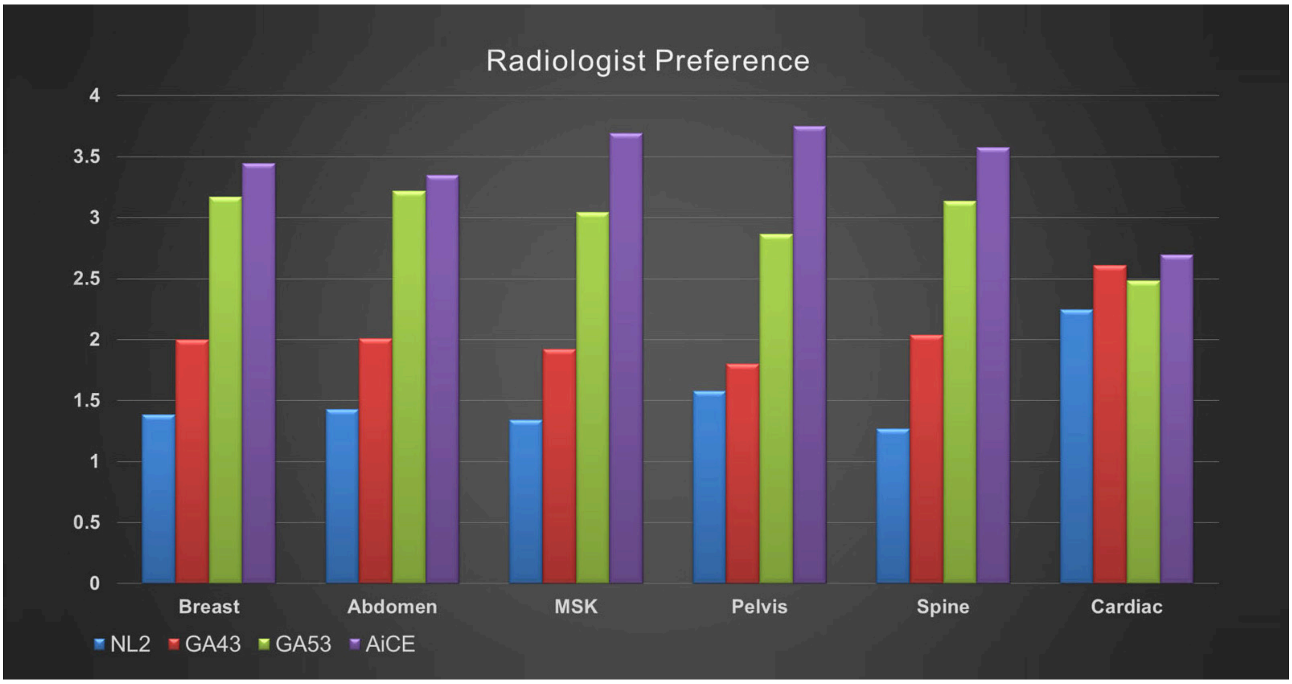

Figure 5 Forced-ranking summary for all anatomies at 1.5T. AiCE is preferred compared to the other reconstructions.

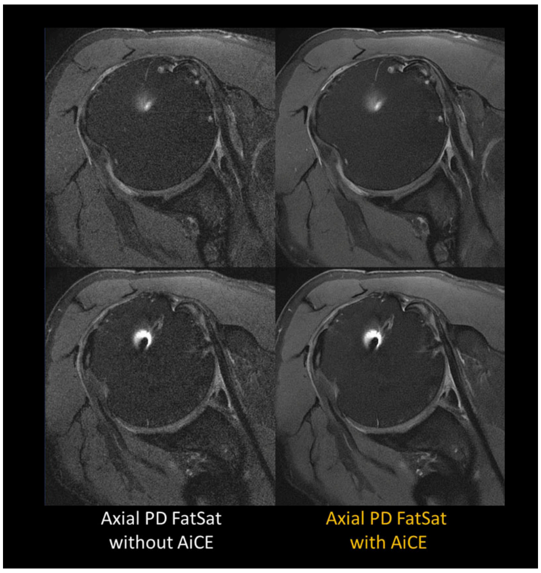

Figure 6 High resolution shoulder without (left) and with (right) AiCE at 3T.

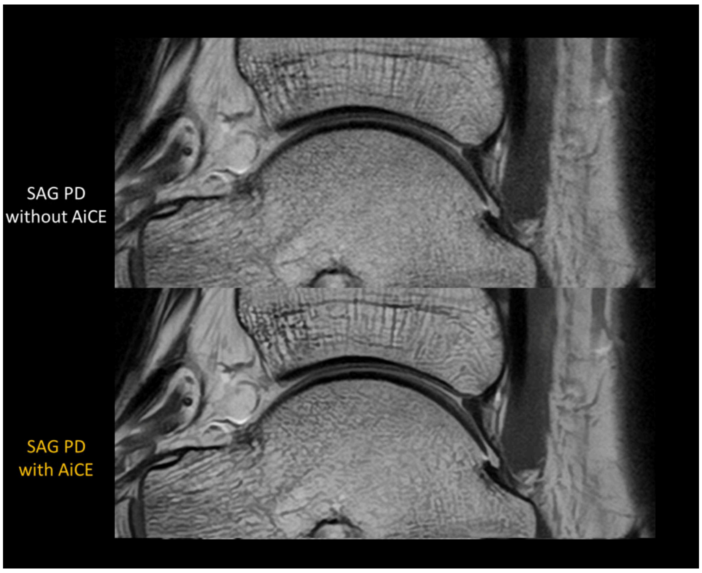

Figure 7 High resolution ankle without (top) and with (bottom) AiCE at 3T.

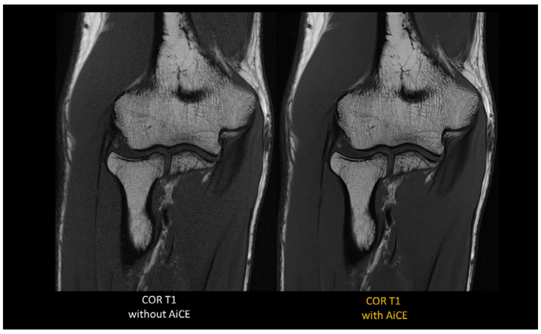

Figure 8 High resolution elbow without (left) and with (right) AiCE at 3T.

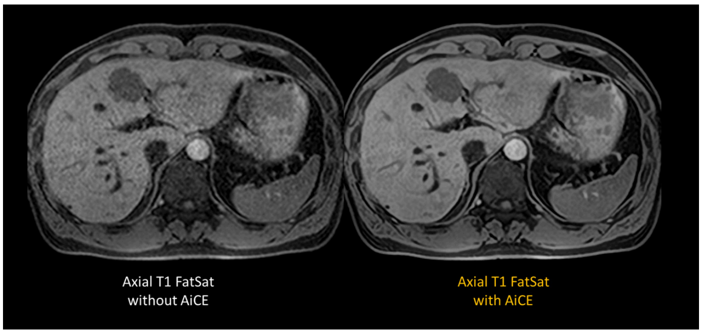

Figure 9 High resolution liver without (left) and with (right) AiCE at 1.5T.

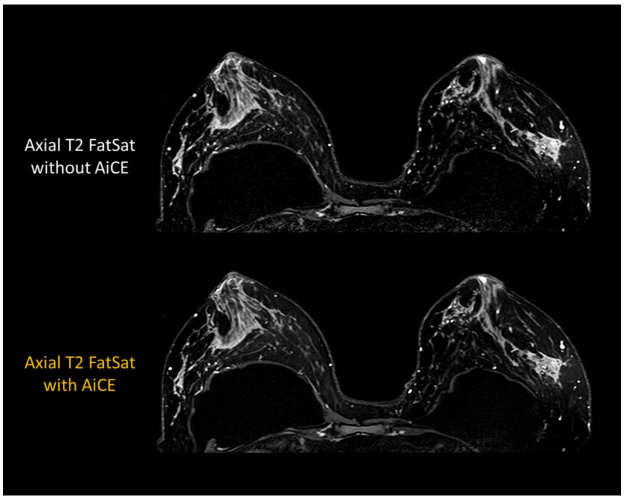

Figure 10 High resolution breast without (top) and with (bottom) AiCE at 1.5T.

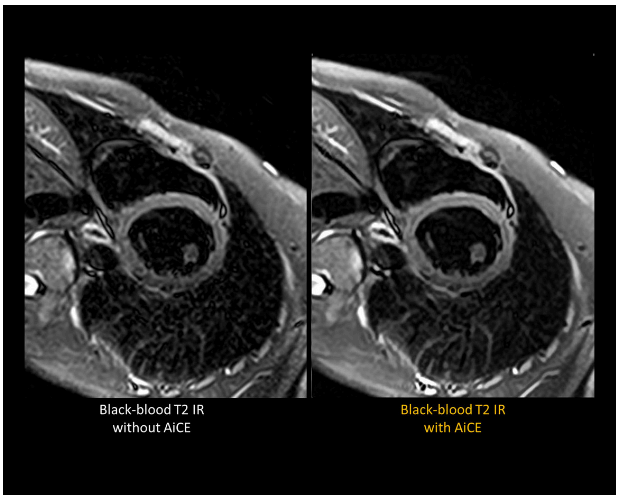

Figure 11 High resolution cardiac without (left) and with (right) AiCE at 1.5T.

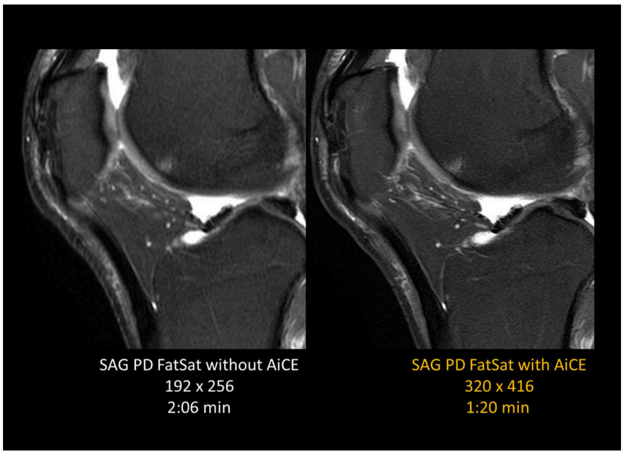

Figure 12 Higher resolution and shorter scan time Knee with AiCE at 3T.

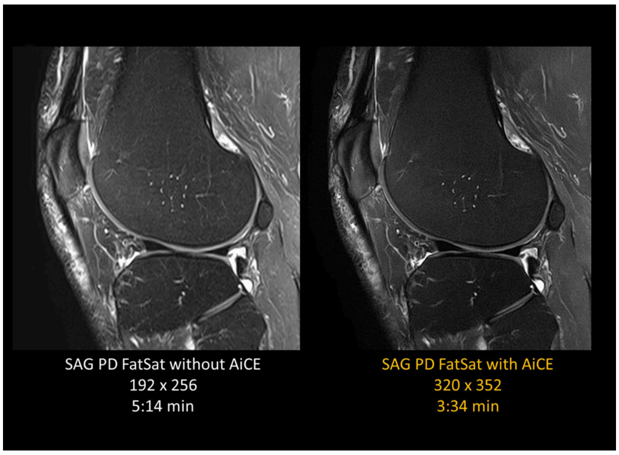

Figure 13 Higher resolution and shorter scan time Knee with AiCE at 1.5T.

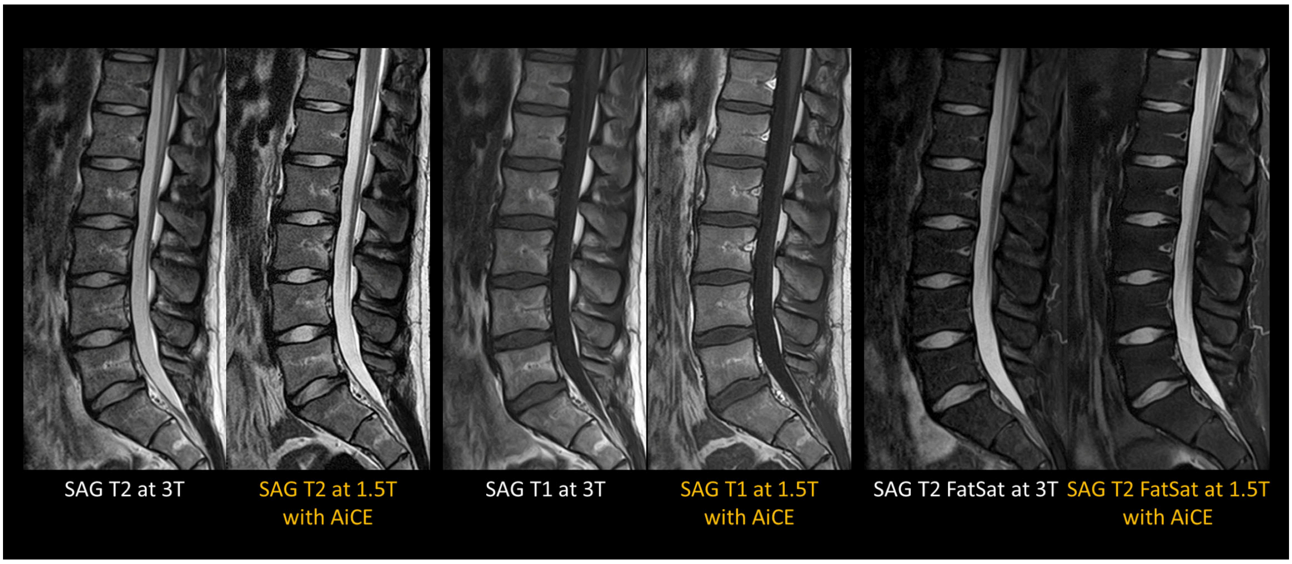

Figure 14 3T images (left) vs. AiCE deep learning based reconstructed 1.5T images (right). The 3T and 1.5T images were acquired with the same resolution and similar scan time while the other parameters were optimized for appropriate contrast at different field strength.

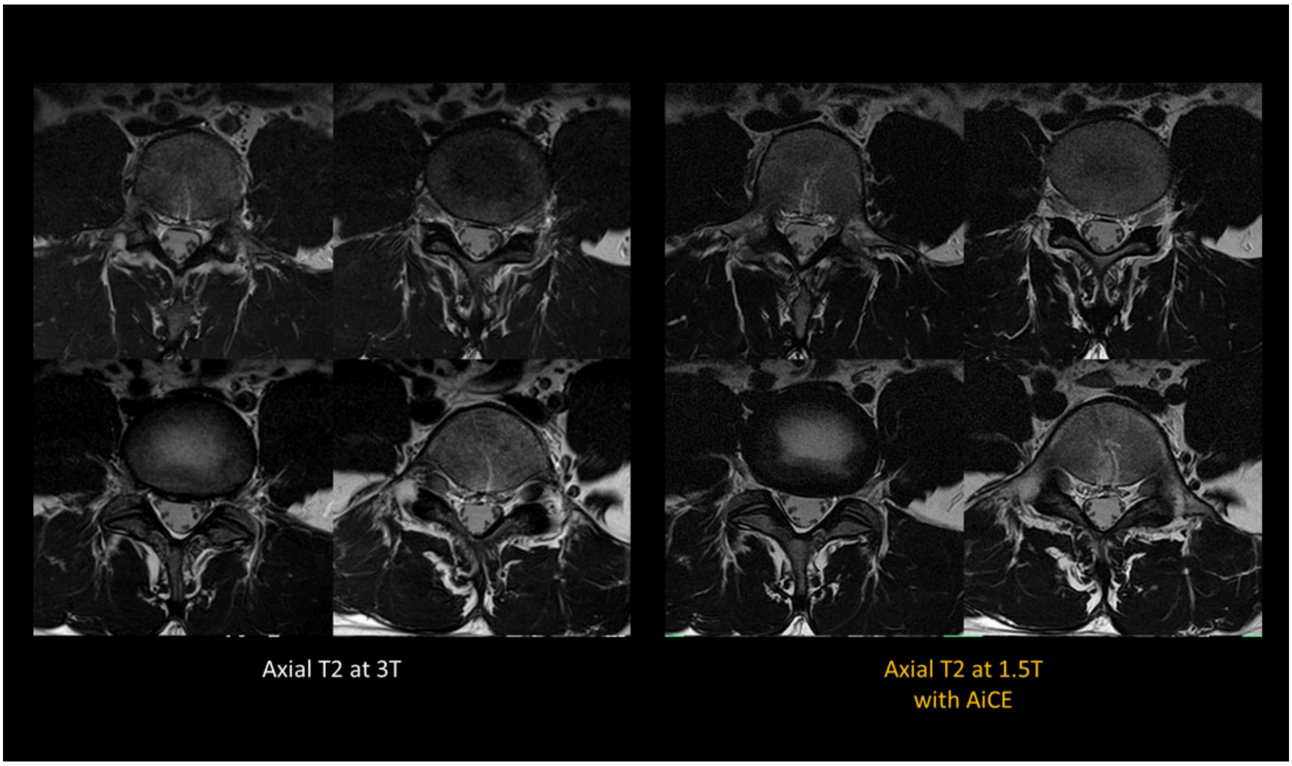

Figure 15 3T images (left) vs. AiCE deep learning based reconstructed 1.5T images (right). The 3T and 1.5T images were acquired with the same resolution and similar scan time while the other parameters were optimized for appropriate contrast at different field strength.

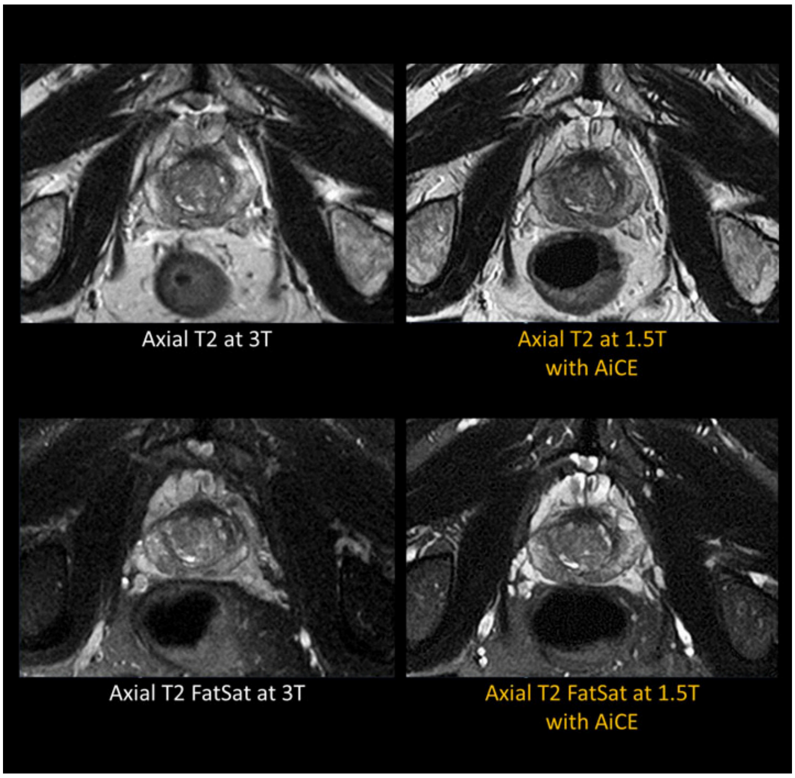

Figure 16 3T images (left column) vs. AiCE deep learning based reconstructed 1.5T images (right column). The 3T and 1.5T images were acquired with the same scan time and resolution while other parameters were optimized for appropriate contrast at different field strength.

Summary

In summary, rigorous validations and the synergy between the domain expertise and the power of deep learning allow AiCE model to be interpretable and its robust and generalized performance to be explainable. AiCE deep learning reconstruction is applicable to all body regions where it may enable (i) acquisition of high quality MRI images, (ii) acquisition with improved resolution and/or shortened scan time, and (iii) acquisition of high-field liked image quality without the challenges associated with high-field system.

Acknowledgments

The authors would like to thank Takashi Tanaka, Kensuke Shinoda, Erin Kelly, and Jonathan Furuyama for reviewing the paper and providing valuable feedback. The authors would like to thank Kirsten Boedeker and Jay Vaishnav for providing useful information related to the MO study.

References:

1. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521:436–44.
2. Benjamens S, Dhunnoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. npj Digit. Med.. 2020;3:1–8.
3. Rezazade Mehrizi MH, van Ooijen P, Homan M. Applications of artificial intelligence (AI) in diagnostic radiology: a technography study. Eur. Radiol. European Radiology; 2020; 1-7
4. Kidoh M, Shinoda K, Kitajima M, Isogawa K, Nambu M, Uetani H, et al. Deep Learning Based Noise Reduction for Brain MR Imaging: Tests on Phantoms and Healthy Volunteers. Magn. Reson. Med. Sci.. 2019;mp.2019-0018.
5. Do HP. Advanced intelligent Clear-IQ Engine (AiCE): Translating the Power of Deep Learning to MR Image Reconstruction. Canon Med. Syst. Corp. 2020;MWPMR0004EA.
6. Do HP. Advanced intelligent Clear-IQ Engine (AiCE) Deep Learning Reconstruction : Effectively Removes Noise while Maintaining MR Signal. Canon Med. Syst. Corp. 2020;MWPMR0007EA.
7. He X, Park S. Model Observers in Medical Imaging Research. Theranostics. 2013;3:774–86.
8. Barrett HH, Yao J, Rolland JP, Myers KJ. Model observers for assessment of image quality. Proc. Natl. Acad. Sci. U. S. A. 1993;90:9758–65.
9. Leng S, Yu L, Zhang Y, Carter R, Toledano AY, McCollough CH. Correlation between model observer and human observer performance in CT imaging when lesion location is uncertain. Med. Phys. 2013;40:0819080-(1-9).
10. Li K, Garrett J, Chen GH. Correlation between human observer performance and model observer performance in differential phase contrast CT. Med. Phys. 2013;40:111905-(1-14).
11. Antun V, Renna F, Poon C, Adcock B, Hansen AC. On instabilities of deep learning in image reconstruction and the potential costs of AI. Proc. Natl. Acad. Sci. 2020;1–8.
12. Cheng, J. Y., Chen, F., Sandino, C., Mardani, M., Pauly, J. M., & Vasanawala, S. S. Compressed sensing: from research to clinical practice with data-driven learning. arXiv preprint arXiv:1903.07824 (2019).

**AiCE provides higher SNR compared to typical low pass filters.*

CANON MEDICAL SYSTEMS USA, INC.

https://us.medical.canon

2441 Michelle Drive, Tustin, CA 92780 | 800.421.1968

MRWP13668US

Made For life