White paper
Nuance Precision Imaging Network

Nuance and NVIDIA: simplifying the translation of trained imaging AI models into deployable clinical applications.

Bridging the gap from medical imaging AI development to clinical adoption quickly, safely, and effectively.

Contents

Nuance and NVIDIA have partnered to bring medical imaging-specific machine learning development frameworks directly into clinical workflows. This will enable the fast, effective validation, deployment, and evaluation of medical imaging AI models to drive better patient outcomes and business performance.

Executive summary

The value and potential of machine learning (ML) for healthcare is no longer theoretical. ML has demonstrated high levels of performance across a wide range of clinical domains and use cases—chiefly in the field of medical imaging.

Between 2014 and 2021, more than $2.7bn was invested into the development of medical imaging AI, with the value of the global AI for medical imaging market set to hit $20.18bn by 2031.[1] But despite the scale of that investment—and the FDA's approval of numerous solutions for clinical use—it still is not translating into widespread clinical adoption of AI for two main reasons.

First, today's healthcare environments are still highly fractured and made up of complex networks of stakeholders.

Secondly, most of today's teams lack standardized, healthcare-specific platforms and integrated development and run-time environments (IDEs and RTEs) for medical imaging AI. Their existing clinical technical infrastructure typically lacks the functionality, established workflows, and governance required to easily validate, deploy, and monitor medical imaging models in real-time. This perhaps poses the biggest barrier to effective clinical adoption of AI.

Now, that is set to change. Recently, NVIDIA, Microsoft, and Nuance partnered to address these universal challenges and accelerate the development and deployment of clinical imaging AI models. By combining MONAI, a domain-specific medical AI platform accelerated by NVIDIA, and the Nuance Precision Imaging Network (PIN) platform, we're bringing medical imaging-specific machine learning development frameworks directly into clinical translation workflows for the first time—effectively accelerating innovation and clinical impact from "bench to bedside".

In this paper, we'll explore how this landmark partnership is helping to close the translation gap and transform workflows for radiologists, other clinicians, and their organizations—enabling accelerated model development, safe and effective validation, deployment, and evaluation of medical imaging AI models across the healthcare landscape.

CONTRIBUTING AUTHORS

1. **Alvin Ihsani,** Deep Learning Solutions Architect, NVIDIA
2. **Brad Genereaux,** Global Lead, Healthcare Alliances, NVIDIA
3. **David Bericat,** MONAI Deploy Working Group Lead, Product Management, NVIDIA
4. **David Werth,** Senior Software Development Manager, Nuance Communications Inc
5. **Iain Henderson,** Principal Software Engineer, R&D, Nuance
6. **Joe Najjar,** Director Diagnostic Applications R&D, Nuance Communications Inc
7. **Ming Melvin Qin,** MONAI Deploy App SDK Architect, Cloud Software, NVIDIA
8. **Shrikant Deshpande,** Senior Director, Product Management, Precision Imaging AI Network, Nuance Communications Inc

Radiology 4.0: desired workflow for providers

For several years, the medical imaging industry has been pushing what is possible through the development of breakthrough AI models designed to support image segmentation, classification, and detection. Researchers and developers collaborate with radiologists and other physicians to create ground truth data and train new algorithms, using tools like MONAI Core on Microsoft Azure ML.

Unfortunately, once their models are ready, that's often where the collaboration ends—there's still a long way to go before AI results can be made available for clinical use, and delivered through the tools radiologists and other clinicians use every day.

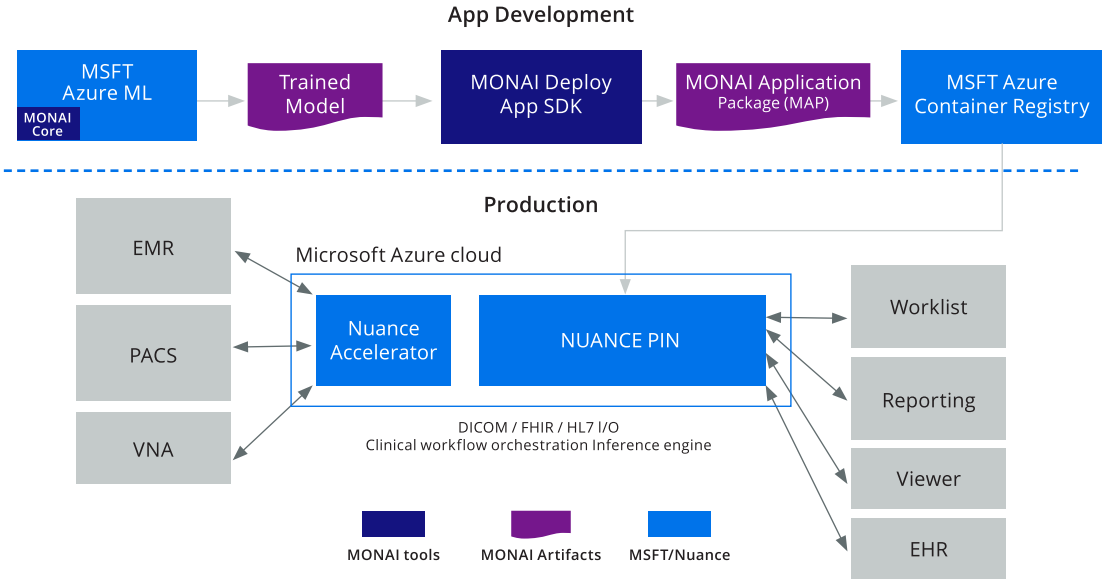
In medical imaging, patient records are typically represented as DICOM studies, which are normally stored in a PACS or VNA. AI inference may be used to calculate, for example, an organ segmentation mask or the probability that a tumor is benign or malignant. Before an AI inference can be performed, the imaging study must be pre-processed, to select the right DICOM series and volumes and to apply the right transformations to them. Once the AI inference has been performed and insights have been identified, some post-processing is still required to create the right interoperable objects, contours, and metadata necessary for a radiologist to accept or reject the results.

These steps are what differentiates a trained AI model from a clinically deployable application. A key goal of Microsoft and Nuance's partnership with NVIDIA is to make the process of translating trained models into deployable clinical applications faster and easier for everyone.

The MONAI Deploy App SDK is a developer toolkit that can be used to create MONAI application packages (MAPs) using just a few lines of code, in as little as a few minutes. These MAPs include one or more AI models in a single container, as shown in the workflow diagram in Figure 1.

Using that portable application container, researchers and data science teams can work with hospital IT operations to deploy them into Nuance Precision Imaging Network (PIN) via the Microsoft Azure Container Registry. Nuance PIN connects with hospital systems like medical devices, DICOM routers, PACS/VNAs or EHRs, to ingest data and execute the right MAPs, sending the results back to the radiologists in their preferred clinical viewers, worklists, or reporting tools, like PowerScribe.

Fig 1: AI application lifecycle from research to clinical production



Turbocharging the AI workflow powers the AI renaissance, with stakeholders standing to benefit

— **Hospitals** can develop and fine-tune AI models and applications that are important to their business and clinical needs, and connect them into their ecosystem as they see fit.

— **Application developers** can focus on their innovative and transformative solutions without being burdened by deployment permutations and last-mile integrations.

— **Validation** can happen collaboratively, where data scientists can develop models and bring them into a familiar clinical environment for clinicians to evaluate their relevance and effectiveness in their own workflow.

The MONAI Deploy App SDK gives developers a framework and the associated tools to design, verify, and analyze the performance of AI-driven healthcare applications. With inference at the heart of the SDK, the SDK enables teams to rapidly prototype and experiment with applications. By using DICOM, HL7 and FHIR, and IHE profiles in a portable and scalable cloud-native containerized package, it makes it easier to deliver new apps into the healthcare ecosystem.

Nuance PIN is a network built on proven diagnostic imaging solutions used in 80% of the US radiology market.[2] With more than 12,000 connected healthcare facilities[3], it integrates AI into clinical workflows to deliver real-time clinical intelligence to multiple imaging stakeholders. It combines the power of proven AI with access to vast image sharing and detailed image reporting to help solve the most pressing issues in patient care. AI solution developers using Nuance PIN gain access to a vast network of healthcare organizations and a seamless way to deploy their AI applications at scale.

Alone, both MONAI Deploy App SDK and the Nuance PIN provide many benefits to radiologists and other imaging stakeholders. Now, by bringing the two together, the full transformational potential of both can be realized.

When MONAI applications are deployed on Nuance PIN:

— **AI application developers** get direct clinical integration into PowerScribe, used by ~80% of all radiologists every day[4]

— **Application developers** can focus on their core product rather than struggling with the deployment infrastructure and lifecycle

— **Resiliency and scale** are handled by the PIN environment in Microsoft Azure, offering various configuration options, customized to the individual subscriber and various clinical stakeholders and workflows

— **Nuance PIN deployment** offers a way to monitor AI application performance and model drift post-deployment using mPower analytics engine where AI outputs are captured and mapped to a standard common data element (CDEs) specified by American College of Radiology (ACR).

— **Nuance PIN** also leverages the Nuance PowerShare image sharing network to access prior images across disparate sites, enabling advanced AI applications that require many forms of prior imaging

— **Developers** can commercialize and deploy their solutions much faster, facilitating solutions for more personalized patient care and enabling better financial outcomes for care providers or endpoint solution providers

The result is a complete enterprise-wide foundation for the creation and deployment of medical imaging AI solutions, and faster, leaner imaging workflows.

Nuance PIN with MONAI can be used by researchers and developers to construct tens of thousands of AI models, depending on the needs of the healthcare enterprise. AI models could focus on different parts of patient reporting and acquisition workflow, including:

— **Generating insights** to consider as part of the triaging process

— **Pre-processing imaging studies** to segment and measure anatomical structures in three-dimensional space

— **Providing insights** to support primary diagnostics as well as potential secondary findings

— **Drawing on previous imaging studies and reports** to calculate disease progression

MAPs can be built using industry-standard interoperable data formats like DICOM, and use common modality images like CT, MRI, x-ray, visible light (photographs), video, digital pathology, and much more. AI models could focus on segmentation, classification, or detection of diseases or conditions. These studies could look across different body regions, including head, chest, abdomen, and extremities. The MONAI Model Zoo contains open-source models trained on public datasets for research purposes, including:

— The lung nodule CT detection model performs 3D detection of lung lesions in CT images. Models like this could deliver insights into a lung reporting workflow.

— The pathology tumor detection model looks at whole slide pathology images and assesses the probability that the slide has a metastatic malignancy. Models like this could provide insights to the reporting pathologist as they draft their report.

— The ventricular short axis 3label model takes an MRI study and provides image annotations identifying the left and right ventricular pool as well as the left ventricular myocardium. Annotations like this can offer insights to the reading cardiologist and provide reference images in the report to the referring physician.

Nuance Precision Imaging Framework, integrated with PowerScribe, drives the run-time post-deployment

Every day, thousands of radiologists across the globe trust PowerScribe to help them automatically interpret and report on findings they see in diagnostic images. It's a tool they're familiar with, and one that's become a core part of radiology workflow.

Nuance PIN maps AI results directly into PowerScribe, enabling automated reporting on AI application outputs using common data elements. For example, PIN could host a Neurology AI application that can automatically provide volumetric quantification of segmented brain structure as well as visualization. PIN then maps volumetric measurements into a structured PowerScribe report, where radiologists can accept or reject the measurements. This workflow replaces the manual tasks of the radiologist of dictating or transcribing measurements, which can be a time-consuming and error-prone process. Data in the report can provide actionable insights to the treating neurologist or surgeons. Since PIN is mapping the AI outputs into standardized Clinical Data Elements (CDE) supported by the American College of Radiology (ACR), the standardized data can be valuable for further research or analytics to monitor ongoing model performance and model drift.

The combination of MONAI, PIN, and PowerScribe provides a direct workflow from AI model ideation and design through to active use in clinical settings.

For healthcare providers and radiologists, this can help close the translation gap, and put high-value AI applications into their hands, which may improve patient outcomes and accelerate accurate imaging workflows.

The rapid route from design to clinical delivery is also a huge win for AI developers that want to commercialize their models and applications. They can sell what they create (pending FDA clearance) directly through an open and widely deployed infrastructure, offering flexible and tailored delivery, service, and pricing plans.

For stakeholders across the healthcare industry, AI industry, and medical imaging field, it's connecting multiple dots to build a clear and intuitive path from powerful models to better decision-making, accelerated discovery, and collective learning.

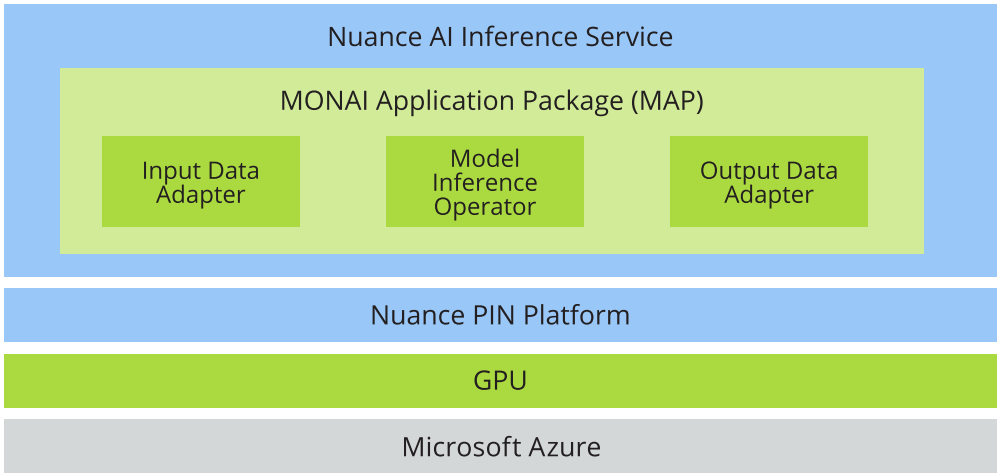


Fig 2: *MAP modular design allows for the decoupled update of data ingestion operators from inference operators and enables the re-development of parts of the MAP from teams with domain-expertise. Application developers update inference components, while IT operations adjust data-interoperability components.*

AI in action at Mass General Brigham

Mass General Brigham (MGB) is one of the academic medical centers closest to bridging the gap between the development of AI models by the research community and their use in clinical practice. MGB uses MONAI Deploy App SDK and Nuance PIN to define a strict "assembly line" that links model development, application packaging, deployment, and clinical feedback for model refinement.

It is not the primary role of researchers to build inference applications used in the clinic, and each of the few inference applications developed is typically bespoke to the model. This results in the loss of intellectual property and an inability to maintain inference applications. Bespoke inference applications are time-consuming and expensive to update (for example, when adopting new data formats for new equipment), and prevent upgrades to libraries, runtime, or the system, to avoid breaking normal functionality.

To overcome these challenges, the data science team at MGB introduced MONAI Deploy App SDK to a dedicated inference application development team. This imposed standardization and modularization of model packaging into an inference application, allowing for updates, upgrades, and modifications to the application that are decoupled from the models (see Fig 2). Intellectual property is packaged in MAPs, made available to the clinical institution for evaluation, and refined and maintained by dedicated teams of developers with domain expertise.

Nuance PIN provides the platform for integrating the results of MAPs with the Nuance suite of tools to allow for the evaluation and examination of inference results. The integration between MONAI and Nuance PIN streamlines the assembly line, from application development to clinical results. It enables MGB to separate concerns between application development and application maintenance. And continuous feedback from the clinic helps improve, adapt, and generalize the model and the application to futureproof them.

— **Model adaptation goes from years to weeks:** The feedback loop from clinic to research accelerates the evolution of mathematical models and saves MGB costs in the clinic by improving throughput. Models that would ordinarily take years to evolve can now improve in weeks due to continuous clinical feedback.[5]

— **Application updates go from months to weeks:** Separation of concerns and domain expertise when futureproofing an application has helped MGB lower development and maintenance costs, as domain experts only need to "tweak" decoupled parts of an application instead of making changes with unknown impact. Adjustments to domain shifts in data (for example, due to changes in demographics, medical equipment, or reconstruction algorithms) now take 3-4 weeks instead of 3-4 months.[6]

"Our AI researchers can focus on developing their models rather than doing all the plumbing underneath. That makes it simpler to get AI-powered insights to our clinicians, so they can provide the best possible care, accelerate time to treatment, and [help our clinicians] improve patient outcomes."

— Dr. Keith J. Dreyer, Chief Data Science Officer, Mass General Brigham

The Nuance PIN platform's support for inference applications cuts costs across MGB's software development and IT teams, replacing ad hoc in-house solutions to detect, diagnose, and repair issues, and reducing the typical 20-40 person-hours to minutes.[7]

MGB's 14,000 AI researchers can now validate and deploy their models directly in clinician's workflows, providing AI-powered insights to support clinical decision-making—with a powerful impact on patient care.

For example, MGB uses a breast density AI model to accelerate read times, reducing the waiting period for mammography results from several days down to just 15 minutes.[8] Women can now talk to a clinician about the results of their scan and discuss next steps before they leave the facility, rather than going through the stress and anxiety of waiting for results.

Now that AI models are easier to deploy in clinical workflows, they're helping MGB improve patient care in other areas too. One model analyzes CT scans to identify adrenal masses that can be difficult to spot on the periphery of images. Another analyzes lung function in COVID-19 patients, looking for indicators that they can be sent home or that their condition may degrade.

With a common platform that makes it simpler to get AI outputs into clinical workflows, MGB is unlocking the potential of AI to transform healthcare.

Learn more

- View a demo
- Nuance Precision Imaging Network or contact us at 1-888-372-1908.
- NVIDIA
- MONAI or try it out today with the MONAI LaunchPad Lab
- Microsoft Azure

ENDNOTES

1 NovaOne Advisor (August 10, 2022). AI In Medical Imaging Market Size to Hit US$ 20.18 Billion by 2030. Credit BioSpace. Retrieved October 25, 2022 from https://www.biospace.com/article/ai-in-medical-imaging-market-size-to-hit-us-20-18-billion-by-2030/
2 Based on Nuance's FY23'Q2 diagnostics penetration report, US only.
3 Based on Nuance's FY23'Q2 diagnostics penetration report, US only.
4 Based on Nuance's FY23'Q2 diagnostics penetration report, US only.
5 Customer remarks from press conference held on 11/10/2023.
6 Customer remarks from press conference held on 11/10/2023.
7 Customer remarks from press conference held on 11/10/2023.
8 Customer remarks from press conference held on 11/10/2023.

About Nuance Communications, Inc.
Nuance Communications is a technology pioneer with market leadership in conversational AI and ambient intelligence. A full-service partner trusted by 77 percent of U.S. hospitals and more than 75 percent of the Fortune 100 companies worldwide, Nuance creates intuitive solutions that amplify people's ability to help others. Nuance is a Microsoft company.

HC_4622 MAR 2023