



The Ohio State University Radiology uses NVIDIA Clara to Deploy AI in a Clinical Workflow

Whitepaper



Table of Contents

Introduction	1
AI Algorithm at OSU.....	2
Challenges of Deploying AI Algorithms.....	4
OSU Solution Walkthrough	5
Compute Engine Selection.....	7
NVIDIA Clara Deploy Software Architecture.....	9
Acknowledgements	10
Project Technical Team	10
Author	11
References	11

List of Figures

Figure 1.	Clinical Workflow	3
Figure 2.	Clara Powered Workflow	5
Figure 3.	NVIDIA DGX Station Components	8

Introduction

Deep Learning is transforming radiology as we know it. Publications on Artificial Intelligence (AI) in diagnostic imaging are packed with Deep Neural Networks (DNN) aimed at augmenting the practice of radiology.

According to Richard D. White, MD, MS (Chair, Department of Radiology, The Ohio State University College of Medicine (OSUCOM)): *"It is crucial for AI researchers to collaborate with expert radiologists and incorporate their insights into the process of developing clinically meaningful AI algorithms."*

The project was spearheaded by Barbaros Selnur Erdal, PhD (Director, Laboratory for Augmented Intelligence in Imaging (LAI²), Department of Radiology, OSUCOM). According to Dr. Erdal: *"It is not a question of whether we want to use AI for clinical diagnostics, but a question of whether we are ready when AI becomes indispensable in the clinical environment."*

Both Dr. White and Dr. Erdal, along with Luciano M. Prevedello, MD, MPH (Vice Chair, Informatics and Augmented Intelligence in Medical Imaging, OSUCOM), have combined their collective diagnostic imaging and engineering expertise to develop the clinical workflow guiding this project.

Deploying AI algorithms in a clinical environment is challenging. The tools and workflows of a data scientist's AI lab are, in general, significantly different from those of radiologists in a clinical environment. In order to bridge this gap, the OSUCOM Department of Radiology, and NVIDIA, partnered to seamlessly scale the deployment of a coronary artery classification algorithm for Computed Tomography (CT) images into the hospital infrastructure. This report describes the clinical workflow, and how The Ohio State University's (OSU) Radiology imaging-research scientists and imaging experts, achieved scalable clinical deployment for their AI-based solution using the [NVIDIA Clara™](#) Deploy SDK.

AI Algorithm at OSU

Using novel data augmentation techniques and transfer learning, the LAI² scientists in OSU's Radiology department, developed an Artificial intelligence (AI) algorithm to classify coronary arteries for the presence vs. absence of atherosclerotic disease on CT images [1]. The images were acquired from patients evaluated in the OSU Emergency Department (ED) for chest discomfort, a symptom suggestive of coronary artery disease. These images were originally classified into two categories by an expert radiologist.

- ▶ Images that showed no signs of atherosclerosis in any of the coronary arteries.
- ▶ Images that displayed signs of mild to severe occlusions in at least one coronary artery.

The annotated data was then used to train the neural network. The results were very promising, with the network showing high classification accuracy. But it needed validation, if it were to be used as a solution for automatic classification of coronary artery occlusion in a clinical workflow.

Figure 1. Clinical Workflow



The challenge for the OSU LAI² science team was to enable expert radiologists to evaluate the AI algorithm in a hospital infrastructure with minimal disruption to their existing workflow.

At OSU, [syngo.via](#) is the integrated imaging software that is utilized in the clinical workflow by radiologists:

- ▶ Once images have been acquired from the scanner, they are sent to the institution's PACS.
- ▶ The patient's study then appears on the radiologist's worklist in PACS, and the images can be opened in Syngo.Via for interpretation.

The AI classification algorithm needs to be seamlessly integrated into this workflow for availability to the radiologist for post-coronary segmentation processing by MeVisLab.

Challenges of Deploying AI Algorithms

- ▶ AI research teams in hospitals generally use varying software tools for developing AI algorithms. Clinical radiologists, in general, are not familiar with these frameworks and software tools.
- ▶ It is important that any evaluation of AI algorithms seamlessly integrates into existing clinical workflows.
- ▶ The AI algorithms that are developed need to be capable of communicating through standard DICOM infrastructure.
- ▶ Crash proof quality-of-service must be guaranteed for the deployment in order to eliminate disruptions to the hospital workflow.



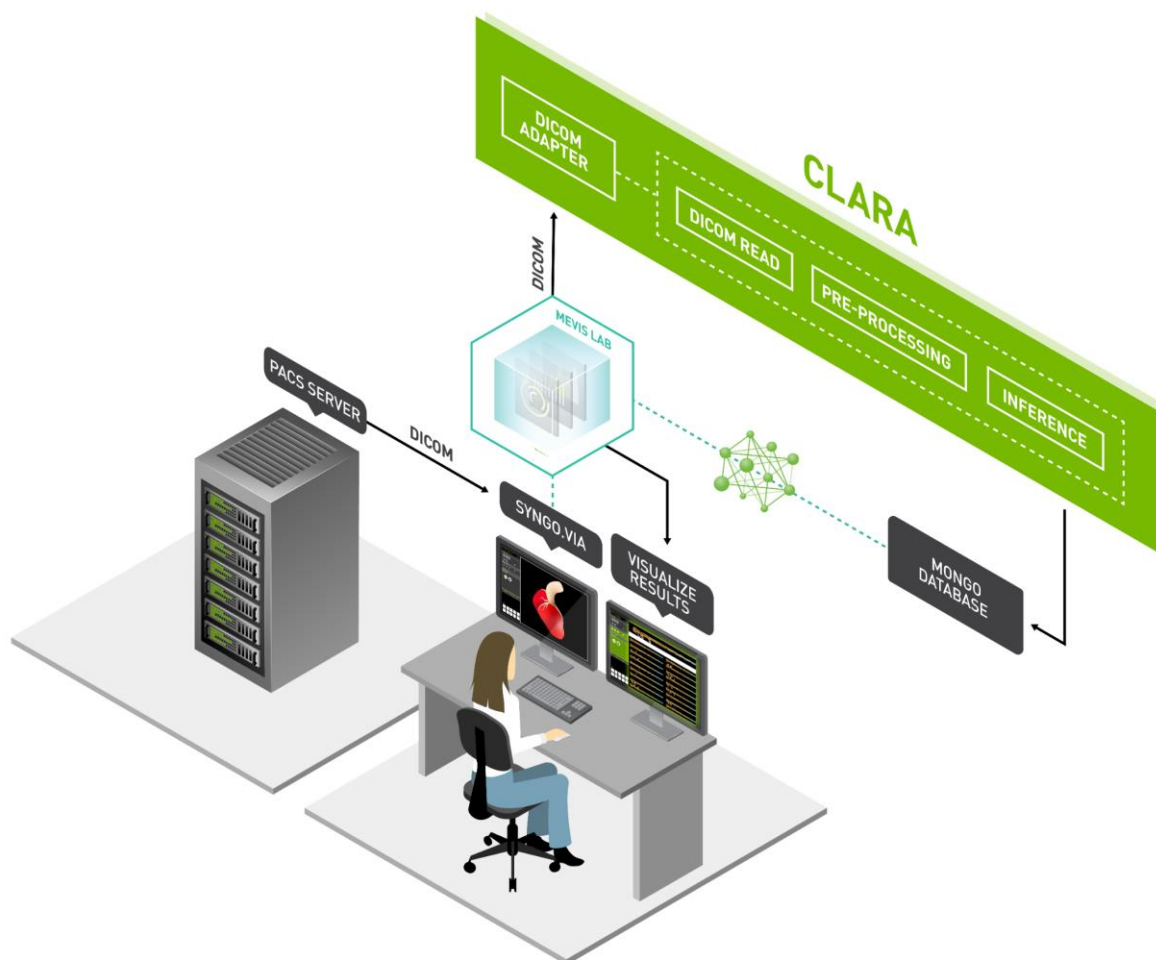
Note: In case of evaluations, the regulatory process is beyond the scope of this whitepaper.

Besides the mentioned constraints on the deployment of AI algorithms in clinical settings, associated costs need to be considered as that may slow the pace of AI research. Hence, there is a need for a cost-effective and scalable way to evaluate and deploy AI algorithms in a hospital setting.

OSU Solution Walkthrough

To take an AI algorithm from a research lab to clinical deployment for evaluation, the OSU LAI² team used [NVIDIA Clara](#) Deploy™ SDK as the underlying framework.

Figure 2. Clara Powered Workflow



The Ohio State University Radiology uses NVIDIA Clara to Deploy AI in a Clinical Workflow
Clara SDKs are for developmental purposes only and cannot be used directly for clinical procedures

Since the goal of this project was to integrate the AI-based coronary classification algorithm into a clinical workflow, it was important for the classification results to be displayed in Syngo.Via. This was achieved using the [NVIDIA Clara Deploy](#) framework and configuring the following workflow:

1. The radiologist opens the patient data in Syngo.Via and selects the most appropriate phase for interpretation. DICOM then sends the series to MeVisLab.
2. MeVisLab processes the input series by segmenting the coronary arteries and creates straight-line Multiplanar Reformations (MPRs) for each vessel. A custom graphical user interface [2] was developed at LAI² for labeling any coronary artery showing signs of atherosclerosis by an expert radiologist (Dr. Richard D. White).



Note: *Segmentation package in MeVisLab is provided by Siemens, and is the same package used inside of Syngo.Via.

3. The straight-line MPR images are sent to NVIDIA Clara Deploy DICOM Adapter.
4. The DICOM Adapter monitors the incoming DICOM images. A classification workflow is initiated every time there is a new DICOM image.
5. The OSU imaging scientists developed a custom DICOM Reader (NVIDIA Clara Deploy SDK includes a reference implementation) that passes the images for conversion from 3D DICOM images into 2D representations for ingestion by the inference-ready classification model.
6. Using NVIDIA Clara Deploy, an NVIDIA [TensorRT](#)[™] inference server is invoked to perform classification inference on the stacked 2D CT images of the coronary vessels.
7. The inference from the classification algorithms produces probability values that are then pushed to a MongoDB database.
8. MeVisLab watches the MongoDB for updated values of the probabilities, and then generates a new DICOM series containing the areas with high probability for the presence of coronary atherosclerosis.
9. This newly created series is then sent to Syngo.Via to be overlaid (fused) on the original images for the interpreting radiologist.

Compute Engine Selection

The OSU LAI² team selected NVIDIA DGX Station™ for its deployment hardware infrastructure. This fully integrated and optimized system enabled the team to quickly start developing a proof-of-concept for the coronary CT Segmentation workflow.

NVIDIA DGX Station features four NVIDIA® Tesla® V100 Tensor Core graphics processing units (GPUs) fully connected through NVIDIA® NVLink®, the NVIDIA high-performance GPU interconnect, delivering 500 teraFLOPS of AI power powered by a complete NVIDIA DGX™ software stack. It includes NVIDIA GPU-optimized containers from NVIDIA GPU Cloud™ (NGC), a software hub for deep learning frameworks, pre-trained AI models and industry solutions such as the [NVIDIA Clara](#) AI toolkit.

The OSU team invested in a hardware infrastructure that could scale with multiple workflows. With Kubernetes-based orchestration, the resources could be partitioned for diverse workloads based on compute requirements.

Figure 3. NVIDIA DGX Station Components

1. GPUs
4X NVIDIA Tesla®V100 32 GB/GPU
500 TFLOPS (Mixed Precision)
20,480 Total NVIDIA CUDA® Cores
2,560 Tensor Cores

2. SYSTEM MEMORY
256 GB RDIMM DDR4

3. GPU INTERCONNECT
NVIDIA NVLink™,
Fully Connected 4-Way

4. STORAGE
Data: 3 x 1.92 TB SSD RAID 0
OS: 1 x 1.92 TB SSD

5. CPU
Intel Xeon E5-2698 v4
2.2 GHz 20-Core

6. NETWORKING
2X 10 GbE

7. DISPLAYS
3X DisplayPort™,
4K Resolution

8. COOLING
Water-Cooled

9. POWER
1500 W

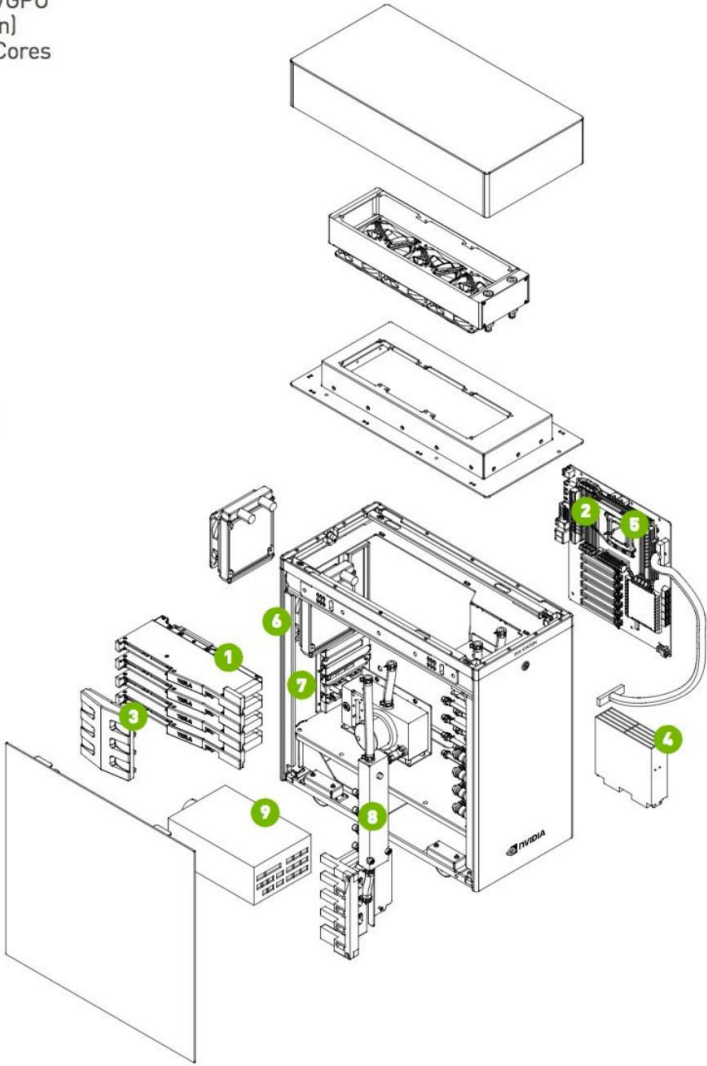


Figure 2 DGX Station™ components

The Ohio State University Radiology uses NVIDIA Clara to Deploy AI in a Clinical Workflow
Clara SDKs are for developmental purposes only and cannot be used directly for clinical procedures

NVIDIA Clara Deploy Software Architecture

The NVIDIA Clara Deploy SDK provides an industry-standard, container-based development and deployment framework for building AI-accelerated medical imaging workflows. The SDK uses Kubernetes “under the hood,” equipping developers and data scientists with the ability to define a multi-staged container-based pipeline.

The main advantage for the OSU LAI² scientists of using a modular framework like the NVIDIA Clara Deploy SDK, was the capability to accelerate the deployment of an AI algorithm in an existing medical imaging workflow. It allowed data scientists to work closely with the expert radiologists and collaborate for faster development of AI-augmented workflows. The key features of NVIDIA Clara Deploy SDK that made this possible were:

- ▶ DICOM Interfacing – NVIDIA Clara Deploy SDK includes a DICOM Adapter that allows receiving and sending of DICOM objects using standard DICOM protocols. The NVIDIA Clara Deploy SDK also includes a DICOM reader and writer reference containers.
- ▶ Crash proof Kubernetes- based deployment which is ideal for hospital environments and scalable for heterogeneous multi-modality workflows.
- ▶ Seamless integration into existing tools through the foundational framework. The data science team at OSU would need to only update the AI base container for the next algorithm that they plan to evaluate.

Here is what Dr. Vikash Gupta, research scientist and the lead engineer for this project had to say about the Compute Engine selection:

“The Clara platform has helped us deploy our AI algorithms and is helping us bridge the gap between AI research and the clinical environment. The engineers at NVIDIA have helped us integrate Clara and address security protocols associated with protecting patient health information, which is a must-have when it comes to deploying clinical solutions. It is great to see NVIDIA acknowledging those strict requirements. Clara AI also supports various integration interfaces and DICOM communication capabilities, which were crucial for tying in different user endpoints such as databases and diagnostic viewers at the Wexner Medical center.”

The Department of Radiology within OSUCOM, under the leadership and direct involvement of Dr. White, translated an AI algorithm that showed high accuracy for the classification of coronary arteries into an integrated clinical workflow that augments the exclusion vs. detection of atherosclerosis in coronary arteries. NVIDIA DGX-Station and NVIDIA Clara Deploy SDK powered the OSU LAI² team with the compute engine foundation necessary to accelerate the development and deployment of AI-powered workflows in medical imaging.

Acknowledgements

Project Technical Team

Dr. Vikash Gupta

Vikash is a research scientist in the Department of Radiology, in the College of Medicine at The Ohio State University. Prior to joining OSU, he was a postdoctoral researcher at Imaging Genetics Center at University of Southern California from June 2015 until February 2018. In March of 2015, he defended his PhD in Computational Biology with a focus on brain imaging and statistics, on the [Asclepios](#) team in [INRIA, France](#). His research interests include statistical analysis, predictive modeling, clustering algorithms and deep learning.

Dr. Mutlu Demirer, OSU

Mutlu received his Ph.D. in 2010 in Electrical and Electronics Engineering from Uludag University in Bursa, Turkey. His research interests include image processing and artificial intelligence.

Matthew Bigelow, OSU

Matt is a certified Nuclear Medicine and CT Technologist with background in imaging informatics. Matt attended The Ohio State University for his Bachelors in Radiologic Sciences and Therapy, and is currently a May 2019 MBA candidate at the Fisher School of Business at The Ohio State University.

Dr. Barbaros Erdal, OSU

Dr. Erdal received his PhD in the field of Electrical and Computer Engineering from The Ohio State University, and serves as an Associate Professor of Radiology and Biomedical Informatics. He is the Assistant Chief of Medical Imaging Informatics at the OSU Wexner Medical Center's Department of Radiology, the Director of Laboratory for Augmented Intelligence in Imaging, and the Director of Scholarly Activities.

Dr. Alvin Ihsani, NVIDIA

Alvin Ihsani is a Deep Learning Solutions Architect researching and developing medical imaging applications and AI infrastructure in a clinical setting. Alvin's background is in applied mathematics in medical imaging for reconstruction, segmentation, and registration, and more recently in deep learning-based image reconstruction and segmentation.

Dr. Ahmed Harouni, NVIDIA

Ahmed Harouni is a Deep Learning Solutions Architect for medical imaging. Ahmed received his PhD in Electrical and Computer Engineering from John Hopkins University focusing on MR image acquisition and image processing. Today, he uses deep learning for image segmentation, registration, disease classification and localization.

Author

Perna Dogra

Perna Dogra is a Product Manager for Healthcare at NVIDIA. Perna drives software products from inception to go-to-markets, and leads medical imaging developer engagements. Prior to healthcare products, Perna managed NVIDIA's Professional Visualization Developer products. Perna holds an MS in Computer Engineering from University of Florida, Gainesville.

References

1. Gupta Vikash, Demirer Mutlu, Bigelow Matthew, Little Kevin, Candemir Sema, Prevedello Luciano, White Richard, O'Donnell Thomas, Wels Michael, Erdal Barbaros. Performance of a Deep Neural Network Algorithm Based on a Small Medical Image Dataset: Incremental Impact of 3D-to-2D Reformation Combined with Novel Data Augmentation, Photometric Conversion, or Transfer Learning, *Journal of Digital Imaging*, 2019 (*in press*).
2. Demirer Mutlu, Candemir Sema, Bigelow Matthew, Yu Sarah, Gupta Vikash, Prevedello Luciano M, White Richard D, Yu Joseph S, Grimmer Rainer, Wels Michael, Wimmer Andreas, Halabi Abdul H, Ihsani Alvin, O'Donnell Thomas P, Erdal, Barbaros S, "A User Interface for Optimizing Radiologist Engagement in Image-Data Curation for Artificial Intelligence," *Radiology: Artificial Intelligence*, Radiological Society of North America

Notice

The information provided in this specification is believed to be accurate and reliable as of the date provided. However, NVIDIA Corporation ("NVIDIA") does not give any representations or warranties, expressed or implied, as to the accuracy or completeness of such information. NVIDIA shall have no liability for the consequences or use of such information or for any infringement of patents or other rights of third parties that may result from its use. This publication supersedes and replaces all other specifications for the product that may have been previously supplied.

NVIDIA reserves the right to make corrections, modifications, enhancements, improvements, and other changes to this specification, at any time and/or to discontinue any product or service without notice. Customer should obtain the latest relevant specification before placing orders and should verify that such information is current and complete.

NVIDIA products are sold subject to the NVIDIA standard terms and conditions of sale supplied at the time of order acknowledgement, unless otherwise agreed in an individual sales agreement signed by authorized representatives of NVIDIA and customer. NVIDIA hereby expressly objects to applying any customer general terms and conditions with regards to the purchase of the NVIDIA product referenced in this specification.

NVIDIA products are not designed, authorized or warranted to be suitable for use in medical, military, aircraft, space or life support equipment, nor in applications where failure or malfunction of the NVIDIA product can reasonably be expected to result in personal injury, death or property or environmental damage. NVIDIA accepts no liability for inclusion and/or use of NVIDIA products in such equipment or applications and therefore such inclusion and/or use is at customer's own risk.

NVIDIA makes no representation or warranty that products based on these specifications will be suitable for any specified use without further testing or modification. Testing of all parameters of each product is not necessarily performed by NVIDIA. It is customer's sole responsibility to ensure the product is suitable and fit for the application planned by customer and to do the necessary testing for the application in order to avoid a default of the application or the product. Weaknesses in customer's product designs may affect the quality and reliability of the NVIDIA product and may result in additional or different conditions and/or requirements beyond those contained in this specification. NVIDIA does not accept any liability related to any default, damage, costs or problem which may be based on or attributable to: (i) the use of the NVIDIA product in any manner that is contrary to this specification, or (ii) customer product designs.

No license, either expressed or implied, is granted under any NVIDIA patent right, copyright, or other NVIDIA intellectual property right under this specification. Information published by NVIDIA regarding third-party products or services does not constitute a license from NVIDIA to use such products or services or a warranty or endorsement thereof. Use of such information may require a license from a third party under the patents or other intellectual property rights of the third party, or a license from NVIDIA under the patents or other intellectual property rights of NVIDIA. Reproduction of information in this specification is permissible only if reproduction is approved by NVIDIA in writing, is reproduced without alteration, and is accompanied by all associated conditions, limitations, and notices.

ALL NVIDIA DESIGN SPECIFICATIONS, REFERENCE BOARDS, FILES, DRAWINGS, DIAGNOSTICS, LISTS, AND OTHER DOCUMENTS (TOGETHER AND SEPARATELY, "MATERIALS") ARE BEING PROVIDED "AS IS." NVIDIA MAKES NO WARRANTIES, EXPRESSED, IMPLIED, STATUTORY, OR OTHERWISE WITH RESPECT TO THE MATERIALS, AND EXPRESSLY DISCLAIMS ALL IMPLIED WARRANTIES OF NONINFRINGEMENT, MERCHANTABILITY, AND FITNESS FOR A PARTICULAR PURPOSE. Notwithstanding any damages that customer might incur for any reason whatsoever, NVIDIA's aggregate and cumulative liability towards customer for the products described herein shall be limited in accordance with the NVIDIA terms and conditions of sale for the product.

Trademarks

NVIDIA, the NVIDIA logo, NVIDIA Clara, NVIDIA Clara Deploy, NVIDIA DGX, NVIDIA DGX Station, NVIDIA GPU Cloud, NVLink, TensorRT, and Tesla are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. Other company and product names may be trademarks of the respective companies with which they are associated.

Copyright

© 2019 NVIDIA Corporation. All rights reserved.