Case Study

2nd Gen Intel® Xeon® Scalable processors
Intel® Deep Learning Boost
Intel® Distribution of OpenVINO™ toolkit
Intel® Distribution for Python*
Medical Imaging

HYHY Creates Full-Cycle AI-Based Medical Imaging Solution to Unleash Potential of Intel® Xeon® platform on AI Acceleration

The implementation of synergistic software-hardware acceleration based on Intel® architecture significantly improved detecting and screening efficiency for diseases such as COVID-19 and breast cancer while also helping medical research platforms avoid the curse of dimensionality

INTEL®
AI BUILDERS
MEMBER

HY 汇医慧影
huiyihuiying.com

"Huiyi Huiying's mission has always been to accelerate the speed of medical imaging analysis through AI technology and help medical institutions complete the transformation from informatization to intellectualization, allowing more patients to benefit from smart medical technology. The introduction of advanced software, hardware products and technologies such as the 2nd Gen Intel® Xeon® Scalable processors, the OpenVINO™ toolkit, and Intel Distribution for Python* has significantly enhanced the processing performance of platforms enabled in our medical imaging solutions."

Chai Xiangfei
CEO & Founder
Huiyi Huiying

The rapid spread of 2019-nCoV acute respiratory disease (hereafter "COVID-19") represented a sudden and very demanding challenge to the rapid diagnostic capabilities of medical institutions. The development of smart analysis and computer-aided diagnosis of medical images with artificial intelligence (AI) techniques offered an effective solution to these new challenges. Huiyi Huiying Medical Technology Co., Ltd. (HYHY) specializes in the application and development of computer vision and deep learning technologies. Using its proprietary Dr. Turing AI platform, Radcloud big data and AI analytics cloud platform, and Novacloud smart imaging cloud product, HYHY supplied medical institutions with a full-cycle, high-performance, and AI-enabled medical imaging solution capable of successfully diagnosing dozens of diseases (including COVID-19).

The Dr. Turing AI platform serves as the linchpin of HYHY's solution and focuses on computer-aided diagnosis of medical images. It provides hospitals and patients with AI analysis of medical images and structured reports for over 10 different types of common diseases. The smart imaging cloud product Novacloud leverages the end-to-end AI interactivity of cloud services to provide doctors, patients, and hospital administrators with more convenient services. Radcloud, the big data and AI analytics cloud platform, serves as an important research tool for medical institutions. Its ability to facilitate scientific analysis of big data from a variety of images greatly accelerates the speed in which new technologies in radiomics and other disciplines can be introduced to the medical industry.

To help optimize the performance of the full-cycle AI medical imaging solution, Intel offered HYHY technologies such as 2nd Gen Intel Xeon Scalable processors with Intel® Deep Learning Boost (Intel® DL Boost) as the core processing engine of this solution, and software optimization tools such as the OpenVINO toolkit and Intel Distribution for Python. Thanks to the advantages brought by synergistic software-hardware acceleration, the Dr. Turing AI platform saw significant improvements to inference speed in image analysis scenarios such as COVID-19 screening and breast cancer detecting. Optimization of the Radcloud big data and AI analytics platform provided significant boosts to the processing speed of the radiomics analysis solution.

The sudden and rapid spread of COVID-19 represented a serious challenge to medical institutions both in China and around the world.

Advantages of HYHY's Full-Cycle AI Medical Imaging Solution

- In pulmonary CT image analysis scenarios related to COVID-19, the combination of strong computing power of 2nd Gen Intel Xeon Scalable processors and optimizations with OpenVINO toolkit reduced the average inference time of the Dr. Turing AI Platform to just 35% of the original platform[1].
- In breast cancer image analysis scenarios, the detection model, constructed on Dr. Turing AI platform based on 2nd Gen Intel Xeon Scalable processors and the INT8 inference acceleration provided by its built-in Intel® DL Boost, gained up to 8.24x improvement in inference speed with a loss of less than 0.17% in accuracy after being converted and optimized with OpenVINO toolkit[2].
- Intel Distribution for Python combined with the computing power of 2nd Gen Intel Xeon Scalable processors increased the inference speed of the radiomic analysis solution based on the Radcloud big data and AI analytics platform by 2.08-2.12 times[3].

The principles for controlling COVID-19 are essentially the same as those for preventing and controlling other infectious diseases — early detection, diagnosis, and treatment. “Early detection”, or “early diagnosis”, refers to when infected individuals are identified from suspected groups in a timely manner. This step is critical and allows for early quarantine and treatment. As shown in Figure 1, Chinese hospitals generally use nucleic acid testing (NAT) as the primary method for diagnosing COVID-19 and computed tomography pulmonary angiography (CTPA) as the secondary method.

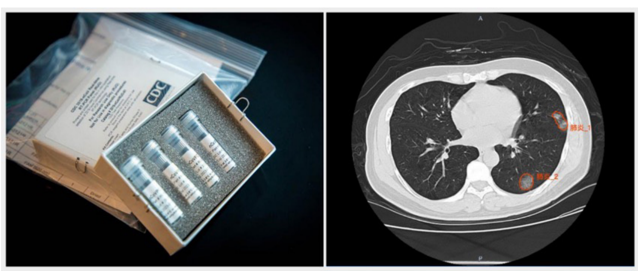

Figure 1 COVID-19 screening and diagnosis methods

NAT is fast and convenient but is affected by factors such as the subject’s viral load, viral mutations, effective extraction of viral RNA, and proper handling of samples. The test has a certain chance of error and may deliver false negatives. CTPA analyzes images of the lungs and searches for pathological changes in the patient. Its sensitivity is superior to that of NAT for early symptoms of COVID-19. One report found that in the testing of 51 patients, the sensitivity of CTPA to COVID-19 was 98%, far higher than the 71% sensitivity of NAT (reverse transcription polymerase chain reaction, RT-PCR)[4].

The application “threshold” for CTPA is far higher than that of NAT, as the physician must be trained in CT image analysis and aware of the clinical signs of COVID-19. For example, Ground-glass opacity (GGO) is one of the most common signs of COVID-19 and an important manifestation for many types of pneumonia and pathogenic infections. Physicians analyzing conventional 2D CT images must often make repeated comparisons to verify their diagnosis. Multi-disciplinary consultation diagnosis may even be necessary in more complex situations. This type of response will obviously lead to delays in pandemic control in the event of an outbreak.

To overcome this problem and to allow CTPA to play a timelier and more important role in pandemic prevention and control, HYHY, a global leader in medical imaging AI applications and research, deployed its Dr. Turing AI platform to provide rapid screening of COVID-19.

Synergistic Software-Hardware Optimizations Accelerate COVID-19 Screening

The Dr. Turing AI platform is part of HYHY’s full-cycle AI medical imaging solution and focuses on AI-aided medical image diagnosis. It offers hospitals and patients two primary functions, "AI-aided applications of multiple diseases" and "structured reporting". In addition to AI-aided diagnosis of more than 10 common diseases, the platform also provides automated lesion segmentation, detecting, and measurement along with structured reports containing **-RADS grading information. The platform improves hospital diagnostic efficiency and helps patients obtain more accurate treatment.

While the Dr. Turing AI platform was a verified and proven solution, it still needed some specific optimizations and upgrades to handle COVID-19 screening. These optimizations and upgrades (shown in Figure 2) include: the platform needed to collect large amounts of COVID-19 data for physicians to tag, then the built-in deep learning algorithms were used to implement precision segmentation and measurement of pneumonia-related lesions. The platform also needed incorporate deep learning algorithm models such as Inception V4 and Inception ResNet V2 to reconstruct two-dimensional CT segments into a three-dimensional model displaying the location, size, change in surface area and severity degree of the lesion in a clearer and more distinct manner. It can provide physicians with more comprehensive pathological image analysis, allowing for more efficient and accurate comprehensive judgments.

Figure 2 COVID-19 screening with HYHY's Dr. Turing AI platform

The above-mentioned optimizations tailored for the CT screening of COVID-19 caused an increase to the amount and complexity of computations required during the execution of inference processes for Dr. Turing AI platform. To remain inference performance at a satisfactory level in such cases, HYHY partnered with Intel to tap into the computing power offered by 2nd Gen Intel Xeon Scalable processors. The OpenVino toolkit was also introduced to overhaul the entire pipeline with the inference engine provided by this toolkit. The optimization tools built in the OpenVINO toolkit were used to convert the existing Nested U-Net and HR-Net network models built with the PyTorch deep learning framework into models, which can take advantages of this toolkit. The benchmark results (Fig.3) show that the entire process took up to 140.3 seconds before optimization, while the inference time for each data sample was just 48.47 seconds after optimization. Inference was improved to only take 35% of its previous time, reflecting a 2.89x increase in performance[5].

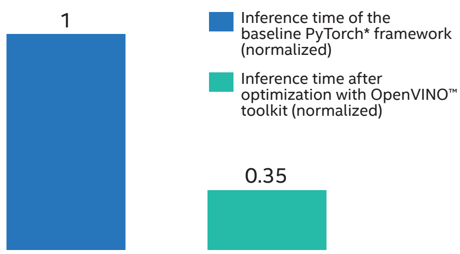

Figure 3 Performance comparison of COVID-19 CT detection model before and after being optimized with OpenVINO toolkit

These improvements in performance allowed for much faster screening of COVID-19 CT images. In fact, this was not the first time that the Dr. Turing AI platform leveraged Intel's synergistic software-hardware optimization to realize robust AI acceleration. HYHY previously adopted the software and hardware combination, which is the core of AI acceleration technology of Intel Xeon Scalable platform, when they took the ResNet50 convolutional neural network (CNN) model as the backbone and developed the RetinaNet target detection model for training and inference on breast cancer screening.

The OpenVINO toolkit includes many types of built-in optimization tools and training models. The toolkit is capable of optimizing trained models through compression, acceleration, and other techniques to enhance inference speed. The optimization method used by HYHY at that time was to optimize the FP32 model for breast cancer detection with this toolkit. The inference speed of the optimized model was 3.02 times faster than the original; HYHY then leveraged the OpenVINO toolkit to convert the above-mentioned FP32 model into INT8 model. Intel DL Boost, built into the 2nd Gen Intel Xeon Scalable processors, was taken advantage of to accelerate INT8 without sacrificing model accuracy. These enhancements increased inference speed by 8.24 times with less than 0.17% decrease in accuracy[6].

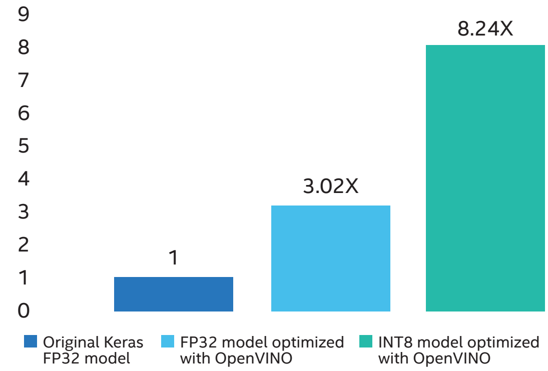

Figure 4 Performance comparison of breast cancer detection model before and after being optimized with OpenVINO™ toolkit

In both the acceleration of COVID-19 CT image screening and optimization of AI breast cancer detection, HYHY's success showed that the combined use of 2nd Gen Intel Xeon Scalable processors and OpenVINO toolkit provided synergy in medical image analysis scenarios greater than the sum of its parts. The success further proved that deploying and operating medical imaging AI applications on the existing CPU-based IT infrastructure present in many medical institutions was not only feasible but also highly effective.

Developing a Full-Cycle AI Medical Imaging Solution

The exceptional performance of the Dr. Turing AI platform in COVID-19 detecting and aided diagnosis represented another milestone in the collaboration among HYHY, Intel, and other partners to help medical institutions realize the critical step of advancing the transformation from informatization to intellectualization through joint innovation.

The rapid development of medical informatics has allowed for widespread adoption of a variety of medical equipment and information systems. Medical imaging devices in particular are being used in an increasing number of applications and have become an important tool for medical institutions and medical research departments, providing critical support for "over 70% of clinical diagnoses."[7] However, only hardware deployment driven by informatization strategy is nowhere near sufficient in the enhancement of the standard of care at medical institutions. Many patients still experience long queues or misinterpretation of medical images, which are remain common in the diagnoses of diseases such as tuberculosis (TB), breast cancer, prostate cancer, and bone fractures where there is a shortage of medical resources.

This is due to the fact that medical image analysis requires not just the right equipment, but also trained physicians proficient in clinical medicine, radiology, and medical imaging. These physicians must not only possess the necessary knowledge and skills, but also be capable of applying many different analytical techniques to realize aided diagnosis. Even large hospitals are experiencing shortage of physicians well trained in all such skills.

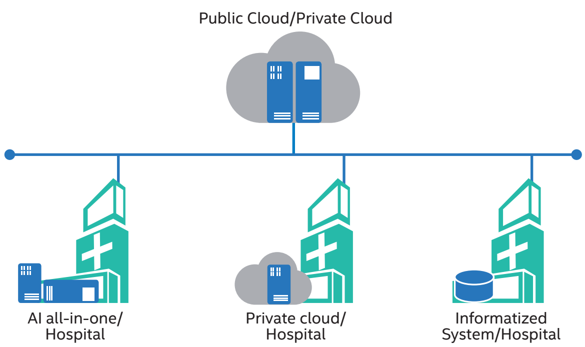

Figure 5 HYHY's AI medical imaging solution designed for deployment in different scenarios

As such, once medical institutions complete deployment of their preliminary information infrastructure, many turn to emerging technologies and techniques such as AI, big data, and cloud computing to realize their intelligent transformation targets. It is for this reason that HYHY developed advanced products such as the Dr. Turing AI platform, Novacloud smart imaging cloud, and Radcloud big data and AI analytics cloud platform. A variety of deployment formats including public cloud, private cloud, or all-in-one device provide medical institutions with a full-cycle, high-performance, and smart AI medical imaging solution.

The Dr. Turing AI platform provides efficient and convenient front-end medical imaging AI analysis and aided diagnosis capability, while the Novacloud smart imaging cloud provides medical institutions and patients with cloud-based end-to-end AI interaction capability. As shown in Figure 6, for physicians, when integrated with the Dr. Turing AI platform, the Novacloud smart imaging cloud provides functions such as multi-terminal AI image viewing, joint telediagnosis, and physician referrals. For patients, the Novacloud smart imaging cloud provides functions such as intelligent report interpretation based on natural language processing (NLP), expert teleconsultation, and multi-terminal image viewing. For medical institutions, the Novacloud smart imaging cloud not only provides support to the Radcloud big data and AI analytics cloud platform, but also enables administrators to configure dashboard that displays diagnosis, treatment, and research progress in a more intuitive manner as well as adjust related policies and strategies as needed.

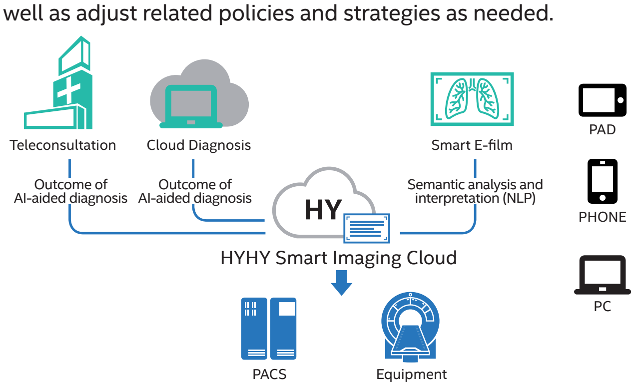

Figure 6 Novacloud smart imaging cloud serving as the final piece of the full-cycle AI medical imaging solution

Intel Distribution for Python Enhances Medical Research Productivity

Apart from the Dr. Turing AI platform and the Novacloud smart imaging cloud, the pillars of HYHY's full-cycle AI medical imaging solution also includes the Radcloud big data and AI analytics cloud platform, which is designed to support medical research.

Currently, many medical institutions utilize medical imaging technologies such as CT, MRI, and PET for the early detection, prevention, and treatment of cancer and other malignant diseases. As a leading developer of radiomics solutions, HYHY is utilizing the Radcloud big data and AI analytics cloud platform as a foundation to provide medical institutions with full-cycle and one-stop big data analytics functionality for medical images to accelerate the development and application of new radiomics technologies.

To better understand the role played by the Radcloud big data and AI analytics cloud platform in radiomics research, we must understand the meaning of radiomics first. Simply put, radiomics refers to process in which medical images are converted into high-throughput image feature data for data mining, after which certain screening techniques are used to extract the radiomics features with the most values. AI model training is then carried out to create a model of values for diagnosis, prognosis, or prediction. The ultimate goal is to provide effective support for precise and personalized treatment.

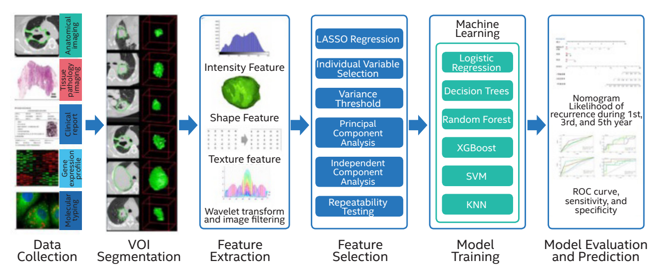

Figure 7 Basic radiomics analytical process

As shown in Figure 7, the basic radiomics process can be divided into data collection, VOI segmentation, feature extraction, feature selection, model training, model evaluation, and model prediction. During the feature extraction phase, the platform will try to extract as many data features as possible to identify details in the medical image. At this stage, the platform may encounter problems caused by the "curse of dimensionality". That is, for a given sample size, as the number of input dimensions increase, the spatial data will become more sparse, exerting a major impact on the predictive performance of the model, greatly increasing the platform's compute load, and reducing inference efficiency.

To overcome this problem, HYHY and Intel chose to utilize feature selection algorithms such as LASSOCV and PCA to effectively shield the research platform against potential curse of dimensionality and help minimize the platform's data loss during compression. To improve the efficiency of these feature algorithms, not only were 2nd Gen Intel Xeon Scalable processors introduced as the computing foundation of the research platform, but the Intel Distribution for Python was also deployed to optimize the overall runtime performance.

Intel Distribution for Python differs from native Python in that it contains the latest vector instructions, supports a greater variety of Intel® libraries such as Intel® MKL and also supports important third-party AI libraries such as sklearn, significantly increasing the platform's execution efficiency during feature selection.

Figure 8 shows the results of the algorithm optimization. The left side shows the LASSOCV algorithm workload when all radiomic features are selected with K-Fold 10 cross-validation. Intel Distribution for Python performed 2.12x faster than native Python. The right side shows the workload of the LASSOCV + PCA algorithm when all radiomic features are selected with K-Fold 10 cross-validation. Intel Distribution for Python performed 2.08x faster than native Python[8].

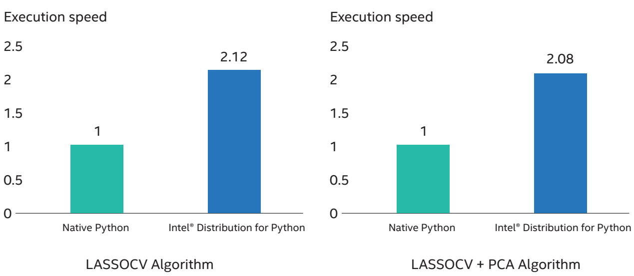

Figure 8 Performance comparison between Intel Distribution for Python and native Python

Supported by the Intel Xeon Scalable platform's synergistic software-hardware acceleration capabilities, HYHY's Radcloud big data and AI analytics cloud platform is currently being used to facilitate research into more than 20 diseases and has successfully been used to apply for over 20 project grants[9]. It is worth noting that the platform's radiomics-based AI solution has been applied to precision identification and diagnosis of COVID-19, and has received good feedback from users in real-world scenarios.

Looking Ahead

The continued implementation and innovation of new technologies such as AI and big data is set to revolutionize the entire medical industry. The introduction of innovative AI medical imaging solutions in particular has led to breakthroughs in many rare and previously incurable diseases. Even infectious diseases, which might be devastating in the past, are now being controlled more effectively with the help of new technologies and platforms.

Looking forward, Intel will continue working with HYHY and other partners to promote application of more advanced products and technologies, especially those related to AI and data analytics, to meet the needs of the medical industry, accelerating the move of medical institutions from informatization to intelligent transformation. Intel and their partners will also work together to develop more diverse and valuable smart medical solutions to benefit the health of the entire population.

HYHY is a subsidiary of Huiying Medical Technology (Beijing) Co., Ltd. and a member of the Intel® AI Builders Program, an ecosystem made up of leading independent software vendors (ISVs), system integrators (SIs), original equipment manufacturers (OEMs), and enterprise end-users in the industry. The vision of this ecosystem is to accelerate AI deployment on Intel® platforms.

[1, 5] Data is based on HYHY's internal testing results using the following configuration: Processor: Dual Socket Intel® Xeon® Gold 6252N, 2.30 GHz; Cores/Threads: 24/48; Operating System: Ubuntu 18.04.4 LTS; PyTorch deep learning framework version: 1.5.1; Intel® Distribution of OpenVINO™ toolkit version: R2020.3.194; Network model: Nested U-Net, HR-Net.
[2, 6] Data is based on HYHY's internal testing results: Https://builders.intel.com/docs/aibuilders/huiying-medical-technology-optimizes-breast-cancer-early-screening-and-diagnosis-with-intel-ai-technologies.pdf. Testing configuration: Processor: Dual Socket Intel® Xeon® Platinum 8268, 2.90GHz; Cores/Threads: 24/48; Intel® Distribution of OpenVINO™ toolkit version: 2019R2; dataset: 366 x-ray mammogram images with an image resolution of 1280x640 provided by HYHY.
[3, 8] Data is based on HYHY internal testing results using the following configuration: Processor: Dual Socket Intel® Xeon® Gold 6252, 2.1 GHz, 24 cores, 48 threads; Memory: 192GB of DRAM; Storage: INTEL SSDSC2BB48; BIOS version: SE5C620.86B.02.01.0009.092820190230; Operating system: 18.04.1 LTS (Kernel: 4.15.0-91-generi); Native Python version: Python2.7.17; Intel® Distribution for Python* version: Intel-Python2019U5; Workload: Classification training on medical images supplied by HYHY.
[4] Data cited from "Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR": https://pubs.rsna.org/doi/10.1148/radiol.2020200432
[7] Data cited from internal statistics maintained by HYHY. For more information, please refer to the article "HYHY Overall AI Solution for Medical Imaging".
[9] Data cited from HYHY's internal document "Application of Radiomics in Medical Imaging and Case Studies".