National Institute of Informatics News ISSN 1883-1966 (Print) ISSN 1884-0817 (Online)

NII Today

National Institute of Informatics News

65
Mar. 2018

NII Interview

[Director General Masaru Kitsuregawa]

Objectives of the Research Center for Medical Bigdata

[Kiyohito Tanaka, Japan Gastroenterological Endoscopy Society (JGES)]

Applying AI Image Recognition in Diagnostic Endoscopy

[Director and Deputy Director in conversation]

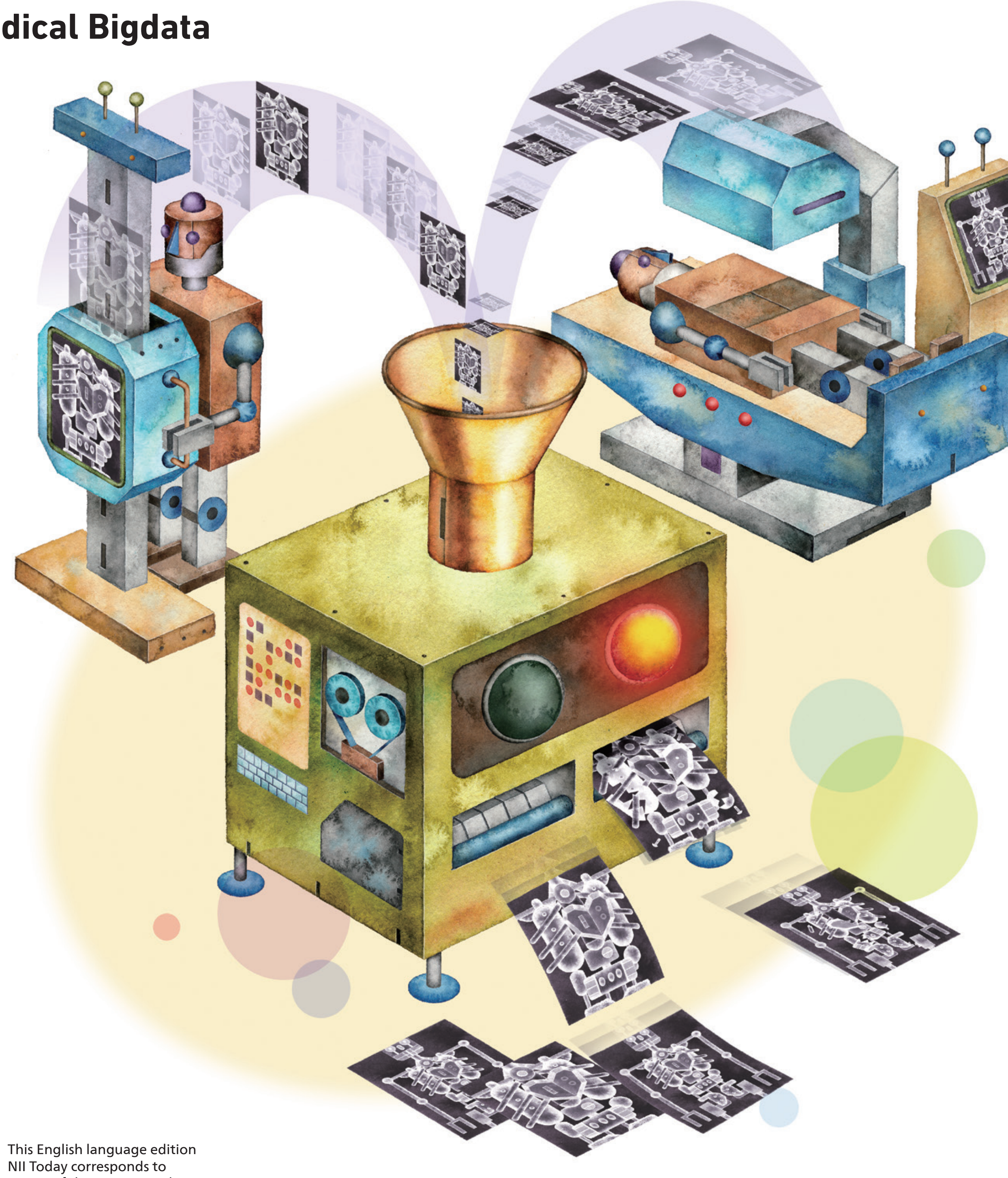
The Power of Image Analysis in the Development of Medical Care

About the Medical Imaging Big-Data Cloud Platform

Feature

New Support for Medical Care Using IT

Launch of the Research Center for Medical Bigdata



This English language edition NII Today corresponds to No. 79 of the Japanese edition

Objectives of the Research Center for Medical Bigdata

Supporting medical care in Japan by collecting and analyzing medical imaging data

Masaru Kitsuregawa Director General, National Institute of Informatics (NII)
Interviewer: **Makiko Tatebayashi** Deputy Editor, Medical News Department, Yomiuri Shimbun

The use of medical big data is gaining momentum in society. In November 2017, NII established the Research Center for Medical Bigdata to promote the collection and analysis of image information from academic societies in health-related fields in Japan. The aim of the Center is to facilitate problem solving in various areas of medicine by using advanced information technology including networking, cloud computing, security, and artificial intelligence (AI). I asked Director General Masaru Kitsuregawa about the purpose of establishing the Center and its future outlook.

Tatebayashi Could you explain the background to establishing the Center?

Kitsuregawa Constructing Information and Communication Technology (ICT) infrastructure for AI and other technologies in medicine and health-related fields is a task within a project called "Realizing the digital revolution in medical care," implemented by the Japan Agency for Medical Research and Development (AMED). So far, four academic societies have been selected by AMED, namely, the Japan Gastroenterological Endoscopy Society, the Japanese Society of Pathology, the Japan Radiological Society, and the Japanese Ophthalmological Society (As of July 2018). NII has partnered with these societies, and this partnership led NII to decide to establish a counterpart.

Each medical society collects and anonymizes medical images from hospitals and other institutions, and transfers them to the "medical imaging big-data cloud platform" built by NII at the end of 2017. Researchers in the fields of Information Technology (IT) and medicine have access to this cloud and can analyze the data collaboratively. The term "AI" is used in the project, but specifically, the research task is to build a system that supports disease diagnosis through medical image analysis using machine learning and deep learning. This means we are aiming to construct a platform that integrates a system of big data circulation and a system of big data analysis.

Tatebayashi So, a hub for medical big data.

Kitsuregawa In medical practice, advancements in examination technology have made it possible to capture vast amounts of information. We are currently entering an era in which a surfeit, rather than a lack, of examination information is presenting problems. Technology aimed at *observing* information will advance in parallel with technology aimed simply at *taking* information. In other words, sophisticated technology that sees and analyzes information and makes diagnoses is becoming necessary. Therefore, we need to develop technology that supports diagnosis.

From the perspective of big data analysis, the process of collecting large amounts of information and turning it into data suitable for analysis is actually very difficult in this field of research. In the current project, we collect information continuously and securely using the super-high-speed Science Information Network, SINET5, built and operated by NII, and we store that information on a secure cloud. When dealing with extremely large-scale data, rather than implementing a conventional style of research in which small amounts of data are collected and small-scale analysis is performed, comprehensive IT strength including networking, cloud computing, and security is essential. Analysis is very important, but analysis alone is wholly insufficient. In the era of data mining, it was often said that the time required for collecting and sorting the data was 90% of the total, while that required for mining was 10% or less, and this has not changed significantly with AI. The success or failure of a project depends on how large an amount and high a quality of data can be collected. This means that as the role of IT in medicine grows, we need to work hard to develop medical big data infrastructure

Masaru Kitsuregawa

directed toward the next stage in Japan.

Tatebayashi How do you think medical care will change with the use of big data?

Kitsuregawa In pathological diagnosis, it is recommended that at least two pathologists make a diagnosis, but I have heard that in reality many hospitals only have one pathologist. IT may be able to provide considerable assistance in uncomplicated areas such as screening.

Even for rare cases that a doctor may come across only once in his or her career, if we have ten thousand endoscopists, we can collect ten thousand cases. Which means that, if we can analyze the collected data and use it as expertise, then we can potentially identify the disease with high probability.

With consecutive medical examination images from a single patient, I think it will also be possible to carry out retrospective research that determines the onset of a disease. In addition, we are also focusing on topics that can only be researched in a country such as Japan, where there is a healthcare system in good practice and every medical institution properly stores examination data.

Tatebayashi What is the appeal of medical big data from an informatics perspective?

Kitsuregawa Today's IT researchers are in great demand. And they are choosing to work on research that needs to be done. We decided that this research was something that NII needed to do. We are using data from an enormous number of people, and maybe there would be resistance from society to this kind of image collection/storage/analysis infrastructure being built by a private company. I feel that perhaps this resistance is reduced slightly when the infrastructure is built by a public research institute. Of course, in fundamental research on image analysis centered on the new field of deep learning, it is very exciting for researchers to implement theoretical research with an eye on applications that will actually benefit society, and it is satisfying to encounter an application topic where sufficient data are available.

Tatebayashi I heard that researchers from institutions other than NII are also participating in this project.

Kitsuregawa That is another major characteristic of this project. From a global perspective, diverse opportunities for competition are available in areas such as image analysis. It is important to create opportunities for various players to meet and compete using their technologies. We currently have participants from the University of Tokyo, Nagoya University, and Kyushu University, and as well as finding ways to get more people from all areas of Japan to participate, we intend to create these kinds of opportunities.

The Institute of Statistical Mathematics (ISM), which like NII, belongs to the Research Organization of Information and Systems (ROIS), will launch the Research Center for Medical and Health Data Science on April 1, 2018. I hope that, in the future, we will cooperate with this ISM Center in certain feasible areas to accelerate research on medical big data analysis.

Tatebayashi What are your thoughts about commercialization of the completed systems?

Kitsuregawa We decided that we are not at the stage of worrying about the intellectual property of things that are not even finished yet, and that we would not think about it at all until the project ends at the end of March 2018. We are also declining all approaches from private companies. However, some medical societies are producing a certain level of results, and so, from April onward, regarding the commercialization, we hope to advance discussions with AMED, from whom we are receiving support. I am involved with a committee connected to the Cabinet Office's headquarters for intellectual property, and I understand that this topic requires in-depth discussions.

Tatebayashi Do you think that Japan could become a world leader in this field?

Kitsuregawa Why is IBM buying medical institutions? Why is Google having autonomous cars drive the equivalent of several times around the globe in tests of these cars on public roads? Because it is impossible to learn using existing algorithms without data, I think that the framework of this project, in which we are working in cooperation with many of Japan's medical societies rather than individual hospitals, did not exist in the past.

Creating this space for big medical image data requires mobilization of technology in diverse information fields. NII has a security team, cloud computing experts, and an infrastructure team, as well as experts familiar with the legal system relating to personal information. Of course, we also have top researchers in image analysis. NII brings together experts from across IT, and I believe that NII's help in mobilizing a comprehensive IT strength makes it possible to offer great flexibility.

However, the project has only just started, so results can be expected from here on out. As we receive appropriate tasks from the medical societies and reliably deliver results, I hope that two-way exchange with the medical side in which they say, for example, "Well, if that's the case, can you solve this problem?" will give rise to a better circulation of information.

It may seem as though I am taking about things that are certain and without risk, but that is not the case. A problem may be so challenging that we cannot solve it, but naturally, we want to attempt to solve outlandish problems too. A lot of fantasies are emerging.

The other day, a doctor complimented us, saying "Wow, you can do that pretty well, can't you!" Our next interview should certainly be interesting!

(Photography by Yusuke Sato)

A Word from the Interviewer

Recently, I have heard doctors at large hospitals express confusion with comments such as, "Companies have approached us with proposals for joint research on AI analysis of clinical data, but we're not sure what to do. We're worried about handling personal information." Even though big data are said to be changing the world, that message does not seem to have gotten through to the doctors busy providing medical care. I think that these doctors will find the framework of this project, which is centered on medical societies and NII, easy to understand.

We now have pioneering technologies such as minimally invasive surgeries and "cancer genome medicine" that analyzes individual patients' genetic makeup to determine treatment. The medical cost is an issue, but the number of patients helped by the use of such advanced technologies is steadily increasing. I have high hopes for frontline efforts to build collaborative relationships in many fields and grow this project so that patients can actually experience the benefits.

Makiko Tatebayashi

Graduated from the Faculty of Science, Kyoto University, and joined the Yomiuri Shimbun in 1988. Worked in the Science News Department before becoming Deputy Editor of the Medical News Department in 2016.

Applying AI Image Recognition in Diagnostic Endoscopy

Aiming to return medical big data to society

Kiyohito Tanaka

Special Advisor to the President, Japan Gastroenterological Endoscopy Society (JGES) Vice Director, Department of Gastroenterology and Chief Information Officer, Kyoto Second Red Cross Hospital

Since 2016, together with NII, the Japan Gastroenterological Endoscopy Society (JGES) has been engaged in Research on the Design of a New, Comprehensive Database Combining a Nationwide Database of Gastroenterological Endoscopy Diagnoses with Endoscopic Images, a topic selected by the Japan Agency for Medical Research and Development (AMED). We asked Kiyohito Tanaka from JGES about the significance of JGES performing this kind of project research, the results obtained so far, and the outlook for the future.

Collecting image information as text

—— Prior to this AMED project, JGES launched the multicenter Japan Endoscopy Database (JED) project in 2015. Please tell us about that project first.

Tanaka The main reason for launching the JED project was to give the information obtained from endoscopy examinations back to the patients. Sixteen million endoscopies are performed every year in Japan. Analyzing these data with a birds-eye view is useful for subsequent diagnosis and treatment, but this kind of analysis is difficult for a single hospital to perform. This is why the society decided to do it.

—— What kinds of data are being collected in JED?

Tanaka In Japan, the results of diagnostic imaging and pathological examinations are often reported in writing, and the terms used may vary. For example, gastric cancer may be written in English as “gastric cancer”, or it may be written in Japanese as “ 胃がん ” or as “ 胃癌 ”.In addition, it may not be clear until the end of the report whether a certain disease is suspected, and this makes database compilation very difficult.

Therefore, in JED, we decided to collect findings as standardized text, and we worked with endoscope manufacturers to develop software that allows doctors to select and enter, for example, a disease name such as “gastric cancer”, or an indication of the degree of atrophy caused by Helicobacter pylori, using a computer mouse. A total of 680,000 cases from nine institutions have already been collected as text data, and the collected data are easy to analyze because there is no variation in the wording.

Rapid progress as a result of meeting NII

—— In the AMED project, images were collected based on JED's text data, right?

Tanaka Approximately 40 images are taken in a typical endoscopic examination. From those images, JED selected several images that clearly showed, for example, gastric cancer, and combined them with the standardized text to create a results report. So we already had tagged image data. We hoped eventually to collect images as well in JED.

However, image data are large and therefore expensive to transport and store, and we had no means of analyzing such data, so we were resigned to the fact that it would be a long time before we could collect image data.That was the situation when AMED suggested that we conduct joint research with NII. In 2016, the research topic was selected and the project began.

—— What kinds of images have you collected and how have you used them?

Tanaka Our aim was to collect images of gastric cancer using text tags, and to use those images to train AI to detect gastric cancer. Images of normal stomachs are also required to diagnose gastric cancer, so those too were collected using tags. We asked NII to do the AI learning part, and because we are using patients' images, we anonymize them all before handing them over to NII.

—— What difficulties have you encountered?

Tanaka Extracting just the images selected for the results reports from the endoscope system was difficult. It seems like it would be an easy task, but it is actually not, because it requires finding a trace of the action “selected.” In addition, the endoscopes were not intended for big data collection, so they only allowed data to be extracted one case at a time. Therefore, the manufacturers worked hard and eventually created software that enables images selected using tags to be automatically extracted all at once.

NII could not start work on the AI learning research while we were waiting for the data collection software to be completed, so we manually collected 14,000 images of gastric cancer from four institutions, marked the lesion sites in 7,000 of these images on a monitor (Figure 1), and used these as “teacher images” to teach AI about lesion sites. We marked up 7,000 images so that there was no deviation from the criteria decided in discussions with NII.

Ensuring results are returned to society by emphasizing an exit strategy

—— What kinds of results are being achieved?

Tanaka At NII, groups from the University of Tokyo, Nagoya

Enhanced-JED Lesion setting application for AI

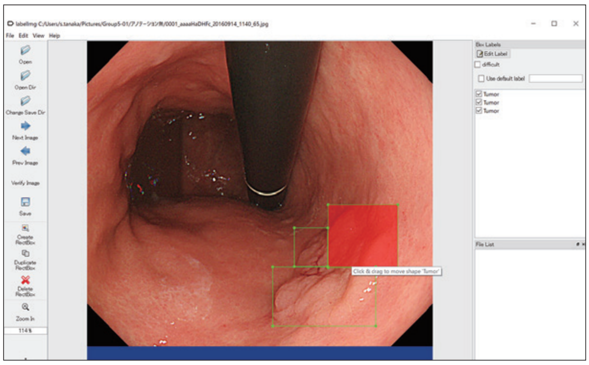


Figure 1 Using the lesion setting application for AI, lesions are marked by enclosing them in a box.

Enhanced-JED Trial of automatic recognition of organ/site

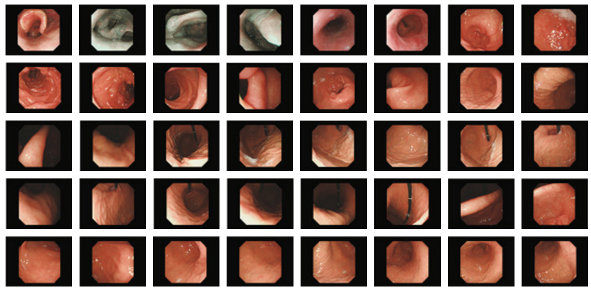


Figure 2 In total, 800,000 images of normal mucous membranes were collected from facilities that had acquired images previously, and these images were analyzed to explore the possibility of automatic recognition/automatic determination of site and organ. If site and organ information can be obtained automatically, it may be useful for accurately evaluating lesions in techniques that are expected to become widely used such as capsule endoscopy, automatic endoscopy, and endoscopic screening.

University, and Kyushu University used different techniques to train AI for images using deep learning, and all of the groups succeeded in training AI to recognize gastric cancer. As they continue to train AI using the collected images, it is expected that it will get even smarter.

Meanwhile, when the AI was given 800,000 images of normal stomachs, it divided them into 60 groups (Figure 2). After looking at all of the images, I realized that these groups corresponded approximately to various sites from the esophagus to the stomach. I was impressed because I had not expected this. I think that this ability of AI will be useful in estimating disease sites from images.

—— Could you tell us about the outlook for the future?

Tanaka While communicating with NII, we have come to understand the problems associated with marking and also the functions required of endoscopes. In addition, gastric cancer is broadly divided into five types and endoscopists diagnose one of those types, but we did not go that far in this project.The project will end at the end of March 2018, and so we are working to get this topic selected again from next fiscal year onward. I hope that we can make AI smart enough to be able to diagnose the type of gastric cancer in more detail, in other words, to make pathological diagnoses. Of course, diagnoses will continue to be made by endoscopists, and some cases will require pathological examination of biopsy specimens, but AI will certainly greatly aid diagnosis.

However, even if we succeed in creating fantastic AI, that will not mean much unless it is actually built into endoscopes and made useful to patients. It is important that manufacturers, the society, and NII work together to advance the research and ensure that it leads to an exit.

—— Are there any requests you would like to make to the Research Center for Medical Bigdata?

Tanaka I think that there are insufficient personnel engaged in medical image AI. A framework that brings many researchers together to work on big data is necessary to develop this field, and I hope that the Research Center for Medical Bigdata serves that purpose. I also hope that the Center makes efforts to return its results to the public and publicize its achievements. Then the patients providing images would really feel that their data are helping to advance medical care.

(Interview/Report by Seiko Aoyama.
Photography by Yusuke Sato)

Kiyohito Tanaka

Graduated from Kyoto Prefectural University of Medicine in 1990, and joined the Department of Gastroenterology, Kyoto Second Red Cross Hospital, in the same year. Has been employed at Kyoto Second Red Cross Hospital ever since, except for one year spent at another hospital. Within gastrointestinal endoscopy, specializes in the pancreatobiliary tract. Is also engaged in medical ICT, including analysis of operating room data, logistics, and the practical use of new electronic technology in hospitals.

The Power of Image Analysis in the Development of Medical Care

Supporting doctors in the field of diagnostic imaging

Tatsuya Harada
Deputy Director, Research Center for Medical Bigdata
Professor, Graduate School of Information Science and Technology, The University of Tokyo
Visiting Professor, National Institute of Informatics (NII)
Team Leader, Machine Intelligence for Medical Engineering Team, RIKEN Center for Advanced Intelligence Project

Shin'ichi Satoh
Director, Research Center for Medical Bigdata
Professor, Digital Content and Media Sciences Research Division, NII
Professor, Graduate School of Information Science and Technology, The University of Tokyo
Visiting Professor, School of Multidisciplinary Sciences, Graduate University for Advanced Studies

The Research Center for Medical Bigdata has started work aimed at developing medical image analysis technology using core technologies of artificial intelligence (AI)—deep learning and image recognition. A joint research project with four academic medical societies is currently underway, but what kinds of research are being carried out and what kinds of challenges have emerged? Also, what kinds of applications are anticipated in the future? The Center's Director Shin'ichi Satoh and Assistant Director Tatsuya Harada are supporting the research as experts in image analysis, and they discussed these questions.

Unanticipated difficulties of medical image analysis

Satoh At the Research Center for Medical Bigdata, we are currently conducting an unprecedented large-scale project to develop medical image analysis technology using *deep learning* and *image recognition,* which are core technologies of AI. This involves collecting medical images from more than 100,000 cases nationwide. The project is being carried out in partnership with the Japan Gastroenterological Endoscopy Society, the Japanese Society of Pathology, the Japan Radiological Society, and the Japanese Ophthalmological Society. Consequently, unlike other NII research centers, the research framework includes external researchers, in addition to in-house experts in networking, cloud computing, security, and personal information. It is reassuring to have a lineup of elite researchers, including Professor Kensaku Mori at Nagoya University, who is achieving results in the development of high-level image processing and its application to medical images, Professor Seiichi Uchida at Kyushu University, who specializes in pattern recognition and machine learning, and Professor Tatsuya Harada, a leader in the field of deep neural networks. What are your impressions after having carried out research for five months?

Shin'ichi Satoh

Harada I have not dealt with medical images in image analysis before, so I have become aware that the difficulties involved are different from those encountered when analyzing general images. For example, the endoscopic images, which I am responsible for, are images of the inside of the esophagus and stomach, so they are soft and squishy and not clearly defined. I have found it difficult to get a computer to recognize these images, unlike items with distinct shapes such as the plastic bottles and cell phones that have been the target of image recognition up until now.

Satoh I was working on analysis technology aimed particularly at television broadcasting, and in the case of television broadcasting or everyday objects, I can judge for myself whether the recognized result is correct. But in the case of medical images, it is frustrating because I cannot judge whether the result is correct.

Training a computer to make a doctor's diagnosis

Satoh One of the aims of medical image analysis is to discover areas suspected of being lesions and subtle differences from normal areas in images. To achieve this, first, we are training a computer to make a doctor's diagnosis, and this requires doctors who are experts in each field to prepare training data by marking areas that support a diagnosis in case images. We are following a process in which the computer is trained using this training data, as well as data from healthy individuals, and the results obtained are discussed by doctors at meetings held once or twice a month. What do you think about this development approach?

Harada We are novices when it comes to medical care, and when we look at pathological images, we do not really know which cells are problematic. It is also essential that we have a two-way exchange with the doctors that includes confirming what level they are looking for, to what extent we can meet their expectations, and whether the technical direction is appropriate. In doing so, we will identify problems with the training data and training model, improve them, and verify the results again. I think that this kind of step-by-step process is the only way forward.

In addition to pathological images, in endoscopic images, which are my assigned area, we have received test data consisting of approximately 100,000 images from healthy individuals and 1,000 to 2,000 case images for training. We are using these data to train the computer, and I think that having too little case data compared to the amount of normal data is perhaps another characteristic of medical image analysis. Even if the amount of case data increases in the future, there will still be an imbalance, so if a smart learning system can be developed under these circumstances, it could be a technical breakthrough. Pursuing algorithms and theories for that purpose is another challenge for researchers.

Satoh The key to increasing the accuracy of image recognition using deep learning is how large an amount of data AI has available to learn from. But, taking pathological images as an example, it apparently takes between a few tens of minutes and one hour to prepare a single image when compiling training data with the basis for diagnosis marked up. The doctors have high hopes for the application of AI in diagnostic imaging, so they are going to great lengths to cooperate despite their busy schedules, and we must respond to that.

Harada The accuracy of image analysis has improved dramatically due to advances in deep learning in the past few years, so I agree that expectations have risen. Unlike the application of AI to the Chinese/Japanese board games *go* and *shogi*, it is difficult to achieve an accuracy that surpasses the top people's skills in this field of research, but if we manage to create a system that can be used to support screening and diagnosis at a certain level, it could be very useful.

Work that researchers find rewarding

Satoh One achievement so far has been in relation to the endoscopic image test data, for which we have developed technology that automatically determines and labels with high accuracy the site where the image was taken, such as the esophagus, top/middle/lower part of the stomach, or duodenum. This will be useful in helping doctors to prepare reports.

Harada In endoscopic examinations, it appears that the doctor takes images at points that should be checked and places of concern while moving the camera from the esophagus to the stomach and duodenum. Which means that the images taken include those about which the doctors did not have any particular feeling and those that the doctor thought looked odd. If one is aiming for early detection, then not only clear cases but also cases where the doctor wavered over their decision are meaningful as data. In the future, it may be necessary to come up with a method of picking up on those cases.

Satoh Cooperation between the different fields of medical care and AI research leads to insights from different perspectives, don't you agree? Hopefully sharing these insights will lead to better results.

Three projects will come to a kind of completion at the end of fiscal year 2017, but the mission before then is to establish a system for secure collection of medical imaging data from multiple hospitals and to conduct pilot studies on image analysis using deep learning technology. With regard to the former, we have created infrastructure for securely transferring and managing large amounts of medical imaging data using SINET5. With regard to the latter, under conditions where only carefully selected data are used for learning, we have achieved a recognition accuracy corresponding to a score of approximately 80 out of a possible 100 points. So, we have had a certain amount of success toward achieving a score of at least 99 points, which is the level required for practical application.

Harada There has been a trend toward applying AI to medical care overseas as well, but given the fact that different races have different characteristic body constitutions and diseases, it is important to develop a diagnosis support system that uses data from Japanese people. Work that leads directly to social benefit, such as advancing medical care, is very rewarding for researchers.

Satoh Much more time is required to realize automatic diagnosis using AI, which is the ultimate objective of research on medical big data, so to start with, we are aiming to surpass the level of a typical doctor in diagnosing common diseases. We will continue working on research that will help to improve the quality of medical care in terms of preventing oversights and improving efficiency by supporting doctors in the field of diagnostic imaging.

(Interview/Report by Akiko Seki.
Photography by Yusuke Sato)

Tatsuya Harada

Completed a PhD at the Department of Mechanical Engineering, Graduate School of Engineering, The University of Tokyo, in 2001. Professor at the Graduate School of Information Science and Technology, The University of Tokyo, from 2013 to the present. Research includes image recognition, machine learning, and intelligent robots.

About the Medical Imaging Big-Data Cloud Platform

Supporting medical image analysis with high security and fast performance

Kento Aida
Professor, Information Systems Architecture Science Research Division,
National Institute of Informatics (NII)
Director, Center for Cloud Research and Development
Professor, School of Multidisciplinary Sciences, Graduate University for Advanced Studies

NII created a *medical imaging big-data cloud platform* to support research and development (R&D) projects on medical image analysis technology. What kinds of technology and functions have been built in to allow secure collection of confidential medical images and enable researchers to smoothly analyze large amounts of data in the cloud? We asked Professor Kento Aida, Director of the Center for Cloud Research and Development, who was in charge of building the platform.

Q What is required for a platform for collecting and analyzing medical imaging big data?

A It has to ensure both *security* and *high performance*. Security is the most important factor for handling highly confidential data such as medical images. The medical imaging data are collected from hospitals by academic medical societies, and after being anonymized by the respective society, are sent to the medical imaging big-data cloud platform. Although anonymized, the data remain medical information, so leaks must absolutely be avoided. High performance of networks and computers is also essential in order to collect huge numbers of medical images that are large in terms of data size from all over the country and analyze them using deep learning techniques, which require computational complexity.

Q What measures have been taken to strengthen security?

A Nowadays, information leaks are a common problem, and in addition to network attacks, many of them are caused by theft or loss with human intervention. Therefore, we decided to place not just the servers but all of the equipment that makes up the cloud platform in a data center with robust physical security measures. This makes it possible to manage the collected medical images securely.

The Science Information Network, SINET5, which is faster than ordinary Internet lines, is used to transmit data from medical society servers and to connect researchers to the cloud. In addition, safety is increased by making it impossible for places that are not pre-configured to connect through the use of a virtual private network (VPN) function (see Page 9), which allow users to connect virtually to a private network.

Q How is high performance being achieved?

A Firstly, SINET5 is an ultra-high-speed network connecting all regions of Japan at data rates of 100 Gbps, so we are able to send data at high speeds. When VPN connections are used with ordinary Internet lines, transmission speeds drop, but an advantage of SINET5 is that its technology allows data to be transmitted quickly and safely.

In addition, processors capable of rapidly executing large numbers of calculations using parallelization and high-performance servers equipped with GPUs are employed to perform deep learning at high speed. Researchers who have actually used the platform have been surprised at how fast the calculations are performed.

Q What were some points that required attention when building the cloud platform?

A In a cloud platform offered to researchers for practical use, it is important to ensure reliability and stability using cutting-edge, proven technology and to provide a forward-looking design. This platform was launched in conjunction with the opening of the Research Center for Medical Bigdata in November 2017, and it has been designed to cope with the demand when the research center is fully up and running and vast numbers of medical images are being sent in. In addition, the file format of medical images varies depending on the device used to take them, but prior interviews held with medical researchers led us to prepare for all file formats.

The work does not end once a cloud platform is built, but rather the platform is improved and developed according to how it is actually used. We will continue to fine-tune this platform while incorporating opinions from researchers in both medicine and informatics.

(Interview/Report by Akiko Seki. Photography by Yusuke Sato)

Kento Aida

Commentary

Tackling Important Issues in the Medical Care Field

Research Center for Medical Bigdata

The Research Center for Medical Bigdata (Director: Professor Shin'ichi Satoh, Digital Content and Media Sciences Research Division), established on November 1, 2017, aims to construct medical imaging big data and develop artificial intelligence (AI) systems centered on image analysis. At the Center, experts in networking, security, cloud infrastructure, and AI technologies, which are fundamental to information systems research, have gathered as the members to be engaged in researching retrieval and recognition of medical images and videos, with a view to developing technology to support diagnosis using deep learning.

To promote this research, first, Professor Kento Aida of the Information Systems Architecture Science Research Division and others are developing and establishing a high-performance cloud platform for medical imaging big data and AI image analysis. This platform will demonstrate its strength in image analysis using cutting-edge AI technology, and will make it possible for researchers to share large volumes of medical imaging data and perform sophisticated image analysis. The development and establishment of this cloud platform will enable medical researchers nationwide to use medical imaging big data easily and securely and to pursue research based on large volumes of data, which was not feasible in the past.

In addition, the cloud platform constructed at the Center is connected via the Science Information Network, SINET5[1], allowing medical care professionals and researchers to access medical imaging data quickly and securely. This will play a major role in the utilization of medical research data.

Meanwhile, joint research with academic medical societies has begun in a nationwide initiative supported by the Japan Agency for Medical Research and Development (AMED). Collecting medical images via the academic medical societies means that images are sent securely to the cloud platform, while image quality is ensured. So far, joint research with the Japan Gastroenterological Endoscopy Society, the Japanese Society of Pathology, the Japan Radiological Society, and the Japanese Ophthalmological Society has begun using medical images collected by each society. From next fiscal year onward, the scale of data collection is due to increase with cooperation from more medical institutions, and research activities will be pursued with a view to developing new links in fields of medical care where problem solving is required.

In analysis of medical imaging big data using AI image analysis technology, the Center is working jointly with the University of Tokyo, Nagoya University, and Kyushu University to put in place R&D systems, establish respective research themes, and attempt to resolve issues. One important topic is constructing a framework for cross-sectoral collaboration together with experts in engineering and informatics, as well as related parties such as doctors and technologists in medical practice. The cloud platform constructed at the Center serves as joint research infrastructure for researchers in medical care and informatics, and will facilitate joint research between the two fields.

Additionally, the Center is exploring the possibility of using the medical artificial intelligence system that will be developed not only in diagnosis support but also in educating doctors. These initiatives can be expected to improve the overall level of medical care in Japan.

Note

[1] SINET5: The Science Information NETwork, built and operated by the National Institute of Informatics (NII). Since its official launch in April 2016, SINET5 has connected all regions of Japan via an ultra-high-speed network offering data rates of 100 Gbps, as well as faster Japan–U.S. channels also offering 100 Gbps and new channels connecting Japan and Europe. More than 850 universities and research institutions across Japan, including all 86 of Japan's national universities, are members of the network.

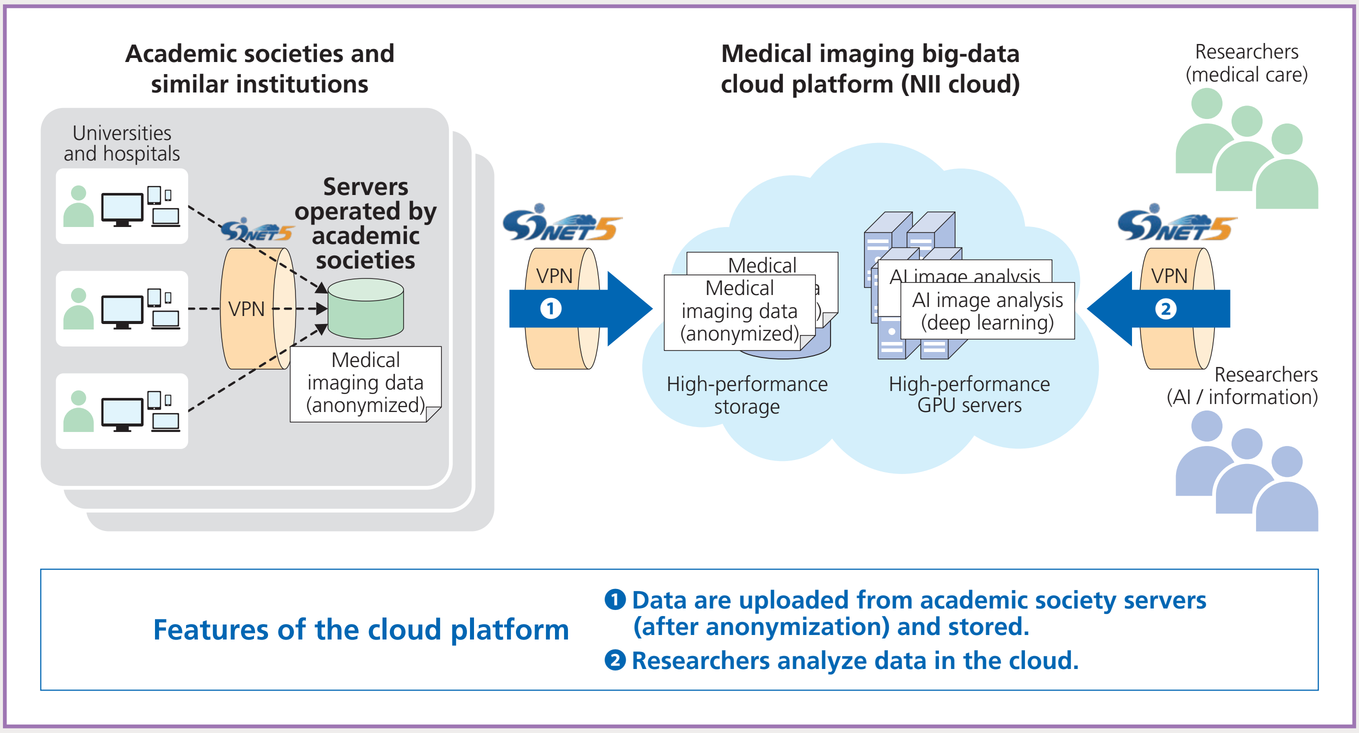


Figure Overview of cloud platform. Medical institutions, universities, and other organizations use the cloud platform via a high-performance virtual private network (VPN) provided by the Science Information Network, SINET5.

NEWS 1 Dreaming new dreams in the world of IT

New Year's speech by Director General Kitsuregawa

Director General Masaru Kitsuregawa, National Institute of Informatics (NII), gave a 2018 New Year's speech to employees on January 9.

He said that NII would continue to actively promote the two pillars of NII—*research* and *service*—and would also combine its expertise in networking, cloud infrastructure, security, and data analysis technologies for the next stage, which is *innovation*. "I want us to dream new dreams in the world of IT."

More than 100 employees gathered at the venue, and it was an opportunity for everyone to reinvigorate their work at NII.

NEWS 2 International exchange agreements (Memorandum of Understanding: MOU) on research with 100 institutions

Agreements concluded with universities and research institutions in 29 countries/regions

The number of institutions with which the National Institute of Informatics (NII) has concluded international exchange agreements (memoranda of understanding: MOUs) on research cooperation reached 100 on January 18 with the conclusion of an MOU with King Abdullah University of Science and Technology (KAUST), Kingdom of Saudi Arabia. NII concludes MOUs with overseas universities and research institutions to promote active and systematic international exchange. The institutions with which NII has concluded MOUs cover twenty-seven countries/regions in Asia, America, Europe, Oceania, and Africa. In addition to MOUs on research cooperation, NII has also concluded MOUs on operational cooperation with nine institutions in five countries/regions.

To encourage the invitation and sending of researchers and students between the institutions with which NII has concluded MOUs, NII offers financial support through the NII International Internship Program and MOU Grants (programs of financial assistance for research exchange), as well as implementing international exchange activities through international joint research projects, seminars, and symposia. With these efforts, NII builds long-term partnerships with the world's leading research institutions and shares research results, thereby improving the quality of research.

NEWS 3 Talk by Emeritus Professor Ullman of Stanford University

NII's invitation taken as an opportunity to host a lecture

Taking the opportunity of inviting Emeritus Professor Jeffrey David Ullman of Stanford University, the National Institute of Informatics (NII) hosted a talk entitled "Data Science: Is it Real?" on February 19.

A leader in database theory research, Emeritus Professor Ullman is also known for writing a large number of books on basic theories of computer science including automata and language theory. He is a recipient of the Special Interest Group on Management of Data (SIGMOD) Edgar F. Codd Innovations Award, the Knuth Prize, the Institute of Electrical and Electronics Engineers (IEEE) John von Neumann Medal, the NEC C&C Foundation Prize, and many other awards.

Following an introduction by Director General Masaru Kitsuregawa, Emeritus Professor Ullman spoke about the various ways in which data science is approached by different communities, including the statistics, machine learning, and database communities (see photograph). He also touched on the education of the large number of data scientists expected to be required in the near future.

Many attendees gathered for this valuable opportunity to listen to Emeritus Professor Ullman speak. The audience listened attentively, and there were so many questions following the lecture that there was not enough time to answer them all.

NEWS 4 SIGVerse to be used as a simulator at World Robot Summit

Developed by Associate Professor Inamura's research group

SIGVerse ™ , a robotic environment simulator developed by the research group of Associate Professor Tetsunari Inamura of the Principles of Informatics Research Division will be used as a simulator for robot competitions at the World Robot Summit (WRS), an international robotics convention sponsored by Japan's Ministry of Economy, Trade and Industry (METI) and the New Energy Industrial Technology Development Organization (NEDO).

SIGVerse is a simulator designed to use robots in virtual-reality spaces, versus actual physical robots, for experiments and assessments involving intelligent robot systems that perform tasks through human–robot interaction. A key feature of this system is that it enables interactions between humans and virtual robots, something that was difficult in previous robot simulators.

SIGVerse realizes real-time communication by combining the Unity platform for VR application development with the Robot Operating System (ROS) software development environment (middleware) for intelligent robots. The system allows human participants to log on to an avatar in virtual space by wearing a head-mounted display, and experience physical and social interactions with virtual robots through body movements, hand movements, and spoken dialogue.

SIGVerse: SocioIntelliGenesis simulator

NEWS 5

Material for teaching information security at higher education institutions

Release of Hikari & Tsubasa's Information Security Class with Three Choice Questions: 2018 Edition

On March 1, the National Institute of Informatics released a PDF version of "Hikari & Tsubasa's Information Security Class with Three Choice Questions: 2018 Edition," which provides material for teaching information security at higher education institutions. The link for downloading the PDF can be found at: http://www.nii.ac.jp/service/sp/.

This released edition is text material (PDF version) containing seventeen episodes of dialogue-style content about information security. The material is structured so that readers will naturally acquire basic knowledge about information security as they follow a story involving four university students as it develops centered around quizzes.

The main target audience is university students who may be experienced in using smartphones and social media but are not sufficiently literate in information security. They can learn about information security by downloading the teaching material to a device such as a personal computer or smartphone, and reading the quizzes and explanations written in each episode.

In addition, teachers at higher education institutions can use this material in introductory lectures on information literacy and information security aimed at first-year students, or as reference material for these lectures. The material can be altered and redistributed for use in lectures provided that the original author is credited (name, title of work, etc.), the material is not used for commercial purposes, and in the case of alteration, the material is published using the same Creative Commons (CC) license combination. Interactive teaching materials will also be released in the future.

NEWS 6

"Multimedia data analysis" and "Online education"

6th & 7th 2017 Public Lectures "The Forefront of Informatics"

On December 18, the National Institute of Informatics (NII) hosted the 6th 2017 Public Lecture series "The Forefront of Informatics." Assistant Professor Yi Yu of the Digital Content and Media Sciences Research Division spoke on the topic of "Knowledge Discovery from the Mass of Images/Videos/Music on the Web—Making Life More Fun Using Multimedia Information" (see left-hand photograph).

Assistant Professor Yu is working on research and development of search/recommendation algorithms and intelligent systems that will offer appropriate support in people's daily lives by analyzing large amounts of multimedia data including videos, music, and text.

As one example, Assistant Professor Yu introduced a video soundtrack recommendation system. The computer calculates geographic and visual characteristics from a video tagged with location information and selects the most appropriate of twenty preset moods, such as *fun*, *angry*, and *quiet*. Next, the computer chooses songs that match the selected mood, calculates the correlation between the audio characteristics of the candidate songs and the user's music listening history, and recommends a song with high correlation to the user. Assistant Professor Yu also explained and demonstrated Event Builder, a system that creates event summary information in real-time, and a (mutual retrieval) system that uses deep learning on correlation between vocals and lyrics to search for lyrics in voice recordings and the reverse search for voice recordings with given lyrics.

The 7th and final lecture of the fiscal year was held on January 30. Assistant Professor Masako Furukawa of the Information and Society Research Division spoke on the topic of "The Potential of Online Education—Using Learning Log Analysis in Learning" (see right-hand photograph).

Assistant Professor Furukawa mentioned the cost to the content provider (such as a university) as an issue in massive open online courses (MOOCs), which have become popular throughout the world, and talked about the need to improve operating costs in order to keep MOOCs "free." As an example of efforts to address this issue, Assistant Professor Furukawa introduced her own research, "Eco-operation framework of MOOC course with previously certificated learners." In 2016, NII offered an introductory programming course titled "Hajimete no P" using gacco, a course providing platform certified by the Japan Massive Open Online Courses Promotion Council (JMOOC), and it offered the same course again in 2017. At that time, twelve people who had completed the course the first time around were asked to act as discussion board supporters, which involved supporting people taking the course by checking the content of posts and commenting on questions. As a result, large disparities were found in the number of comments made by each supporter, and Assistant Professor Furukawa said that she hopes to explore how best to collaborate with learners who have already completed a course.

SNS

"Hey, this is great!" Hottest articles on Facebook and Twitter (December 2017–February 2018)

National Institute of Informatics, NII (official) Facebook
www.facebook.com/jouhouken/

7th Shonan Meeting Special Lecture "Artificial Intelligence Research and Application—Present and Future" (12/26/2017)

National Institute of Informatics, NII (official) Twitter
@jouhouken

Large expansion of Dataset of Japanese Classical Documents. Now available are many cookbooks, the Tales of Ise, books of heraldry, and picture books, which allow readers to enjoy Japanese culture. (12/26/2017)

Bit on Twitter! Twitter
@NII_Bit

Bit says goodbye to his personal designer. Thanks for all the many clothes!! (02/28/2018)

* Some text edited/omitted.

Essay

An Encouragement of Diverse Learning

Ken Satoh
Professor, Principles of Informatics Research Division, National Institute of Informatics

For the past decade or so, I have chosen to research law as a field of application of AI. In this essay, I would like to discuss appropriate approaches to interdisciplinary research based on my experience.

In academia, the development of an isolated, inward-looking mentality is problematic, and the integration of different fields has long been recommended. Attempts have been made to integrate informatics with other disciplines, but success has been limited to areas such as bioinformatics. Common patterns of joint research are Party A working for Party B to solve something that Party B regards as a problem, or Party A having Party B come up with applications for Party A's research.

However, my own experience leads me to believe that true success is not easily achieved using either of these approaches. With the first approach, if Party B is the very best in their field, then research that gets to the root of a problem is possible. But it is not always clear whether Party B is at that level, and therefore, even if Party A solves the problem, the solution might not have a great deal of impact. With the second approach, a solution is presented to Party B without knowing whether it is the optimal method of solving their problem, so it is unlikely to lead to real progress in their field. I think that these problems ultimately arise from a lack of deep understanding of the other party's field, and precious opportunities to gain an understanding of other fields from the new perspective of informatics are being lost using the above approaches to interdisciplinary research.

With these past failures of interdisciplinary research in mind, I used NII's training program to enroll in the law school of the University of Tokyo, with the idea of immersing myself in the field of law for the purpose of applying AI in this field. At first, the enormous differences in vocabulary and concepts led me to frequently question whether AI could contribute anything to law. Fortunately, however, the Presupposed Ultimate Fact Theory, which comprises judicial criteria for reaching judgments using incomplete information and was taught only at The Legal Training and Research Institute of Japan (where successful bar exam candidates gain practical experience for approximately one year) began to be taught at the School of Law when judicial reform took place in Japan.

This theory coincides exactly with the logical reasoning under incomplete information ("nonmonotonic reasoning" in AI terminology) that I was researching as a specialization in AI, and I knew that it could be implemented by computer. I am now building an automatic reasoning system that makes highly reliable judgments regarding civil contract law based on this implementation. This is an opportunity that I would never have encountered had I not entered law school, and I think that it can be regarded as a successful case. Another outcome of this opportunity was that I successfully passed the bar exam in 2015.

The objective of the Inter-University Research Institute Corporation Research Organization of Information and Systems (ROIS), to which NII belongs, is integrative research that goes beyond the boundaries of traditional disciplines by applying the perspective of information and systems to complex issues such as life, earth, the natural environment, and human society that greatly impact our lives in the twenty-first century. In light of this objective, I think that there need to be more successful cases such as the one I have described here.

Both research institutes dealing with information (the National Institute of Informatics and the Institute of Statistical Mathematics) and research institutes dealing with systems (the National Institute of Genetics and the National Institute of Polar Research) exist within ROIS, and there is a wealth of opportunities for integration. Furthermore, if a framework were constructed for mutually recommending researchers to the doctoral programs of the Graduate University for Advanced Studies (for example, tuition fee support), in which ROIS is involved, then I think it would make the deep understanding between fields that I previously mentioned possible and allow true integration. I hope that ROIS headquarters will consider this also in terms of developing the identity of ROIS.

Future Schedule

June 18–19 | Japan Open Science Summit 2018 at Hitotsubashi Hall and other venues. For details and to apply for the Summit, please visit the URL below. https://joss.rcos.nii.ac.jp/

June 20–21 | Academic Information Infrastructure Open Forum 2018, National Institute of Informatics at Hitotsubashi Hall and other venues.

June 22 | Shonan Meetings 100th Commemorative Symposium

June 22–23 | NII Open House 2018 (public exhibition/presentation of research results) at Hitotsubashi Hall and other venues. For details and to apply for events that require advance registration, please visit the URL below. http://www.nii.ac.jp/openhouse/

Notes on cover illustration

This illustration shows robots taking X-rays in a medical checkup and using a machine to determine automatically whether the X-ray images indicate a health problem. I hope that you can find the differences between the two types of data, like in a game of spot the difference.

Bit (NII Character)

Weaving Information into Knowledge
NII

National Institute of Informatics News [NII Today]
No. 65 March 2018
[This English language edition NII Today corresponds to No. 79 of the Japanese edition.]
Published by National Institute of Informatics, Research Organization of Information and Systems
Address | National Center of Sciences 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430
Publisher | Masaru Kitsuregawa **Editorial Supervisor** | Ichiro Satoh
Cover illustration | Toshiya Shirotani **Copy Editor** | Madoka Tainaka
Production | MATZDA OFFICE CO., LTD., Sci-Tech Communications Inc.
Contact | Publicity Team, Planning Division, General Affairs Department
TEL | +81-3-4212-2164 **FAX** | +81-3-4212-2150 **E-mail** | kouhou@nii.ac.jp
http://www.nii.ac.jp/en/about/publications/today/