Methods for Promoting AI and Technological Developments for the Realization

Kiminori Kurosawa,[1] Takahiro Nitawaki,[1] and Tran Duc Son[1]

While many applications of Artificial Intelligence (AI), mainly deep learning have been progressing, they often get stuck at Proof of Concept (PoC) step in manufacturing processes. To efficiently introduce AI into the manufacturing process, we have defined our AI roadmap. Based on this, we are studying, developing, implementing, and operating AI systems at our company. This paper will explain how we are putting effort into AI, details of specific projects, and the technologies we have developed to implement them. We will also discuss a direction of AI system architecture based on new technology movements such as 5G.

1. Introduction

In recent years, AI has been increasingly applied to businesses, and specific effects have been confirmed. On the other hand, the introduction of AI into the manufacturing industry, especially the manufacturing process, has not progressed and often gets stuck at the PoC stage.

We started AI research in 2015 and developed technologies for applying AI to manufacturing processes at our company. Furthermore, since 2018, we have begun introducing AI into the manufacturing processes and are gradually expanding the scope of our application.

In this paper, we report the concept and examples of our works for AI and the technologies we have developed/introduced based on the trends of AI technology. In addition, we will also consider the direction of AI system architecture, taking into account new technology movements such as 5G.

2. Trends in AI Technology

The gap between excessive expectations and reality/ability caused boom and winter eras in general. As for AI, the boom and winter eras have been repeated so far (Figure 1).

The first AI boom began in the 1950s, about a decade after the birth of computers, and continued through the 1970s.

In the first AI boom, most of the theories that led to the current AI technology were already proposed. On the other hand, the implementation technology for computers and the performance of computers themselves were significantly insufficient and could implement only simple games. As a result, the ability could not keep up with expectations for AI, and the first AI boom came to an end.

The expansion of computer languages such as Lisp and Prolog, which have features that make them easy to implement in AI, the construction of knowledge bases for

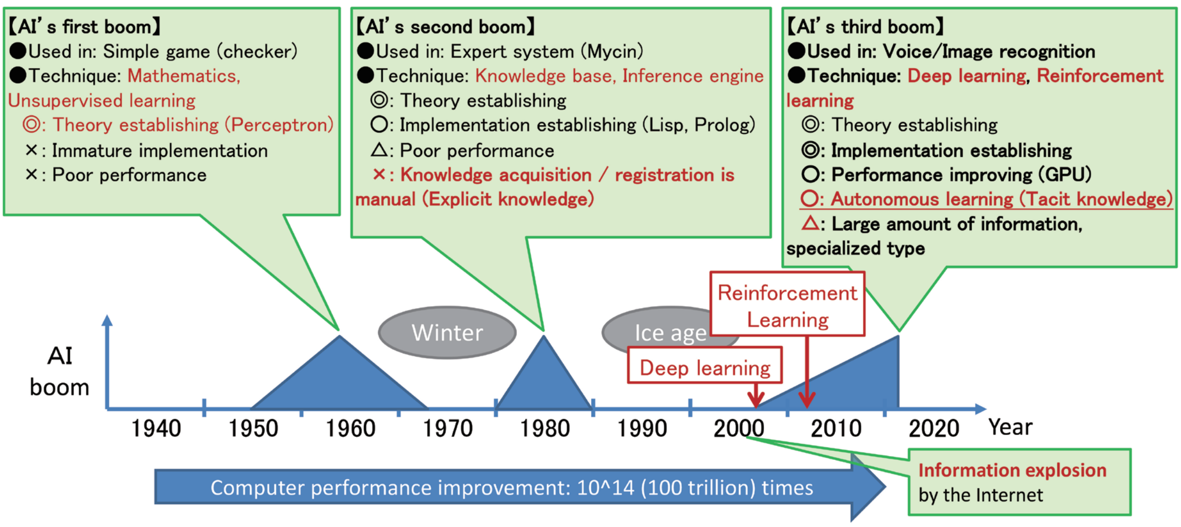


Fig.1. History of AI and the mainstream of third boom.

1 AI Promotion Group, Innovative Fujikura Production System Promotion Department

Panel 1. Abbreviations, Acronyms, and Terms.

AI—Artificial Intelligence
Lisp—A programming language created as a practical mathematical notation for computer programs and preferred in artificial intelligence research
Prolog—A logic programming language associated with artificial intelligence and computational linguistics
Expert System—A computer system emulating the decision-making ability of a human expert
Mycin—An expert systems used in the medical field
Knowledge Base—Database for knowledge management
Deep Learning—A class of machine learning methods that use multilayer neural networks
Reinforcement Learning—A class of machine learning methods that deal with the problem of determining what action to take given the current situation
PoC—Proof of Concept
CNN—Convolutional Neural Network: a class of deep learning networks used for image and video recognition
Object detection—The technology for detecting specific objects in digital images and videos
RNN—Recurrent Neural Network: a class of deep learning used for time series data
Multimodal Deep Learning—A class of deep learning that learns by combining different data types, such as images, audio, and text.
Sim2Real—Simulated to Real: a class of machine learning method that uses simulations to solve real-world problems
Distillation—The process of transferring knowledge from a large model to a smaller one
Gateway—A device or software that relays data between different networks.
FA Network—Factory Automation Network: a network to control equipment in the factory
IP Network—Internet Protocol Network: computer network that uses IP to communicate
Epoch—The number of passes of the entire training dataset the machine learning algorithm has completed
Edge Artificial Intelligence—The system that processes AI algorithms locally on users' devices
Recoat Fiber—Optical fiber that has been recoated after removing the coating
Classification—Machine learning tasks to classify data such as images into a finite number of categories
Transformer—A class of deep learning methods that is mainly used in natural language processing, but research has progressed in recent years and has been applied to fields other than natural language processing.
UPS—Uninterruptible Power Supply: A device that supplies stable power

storing knowledge, and the implementation of inference engines for handling knowledge formed the second AI boom in the 1980s. A typical example is an expert system that makes decisions on behalf of experts. But on the other hand, humans still had to manually perform many works such as explicit knowledge acquisition and knowledge registration in the knowledge base. Furthermore, a lot of effort, such as adding new knowledge and ensuring consistency between them, is required continuously, making it difficult to maintain the knowledge base, and the second AI boom has ended. In short, the reason for the end of the second AI boom was the total reliance on humans for knowledge acquisition.

In the 2000s, the development of the Internet led to an explosive increase in the volume and accumulation of information. Furthermore, computer performance has improved dramatically. Based on this background, deep learning and reinforcement learning that enable autonomous learning of knowledge by AI itself have been implemented. These are the mainstream of the third AI boom. On the other hand, the knowledge acquired through deep learning and reinforcement learning is a black box for humans because it is tacit knowledge that only AI can understand. In addition, there are problems such as the need for a large amount of accurate information to acquire knowledge and specialization for specific purposes, and further evolution is required.

Although there are various AI technologies, we are working on deep learning and reinforcement learning, which are the mainstream of the third AI boom. All the AI described in the following sections refers to deep learning or reinforcement learning.

Chapter 3 describes the concept of the approach to AI, chapter 4 describes specific examples of the efforts and introduced technologies, and chapter 5 describes the AI system architecture.

3. The Concept of the Approach to AI

In general, AI systems are often discussed by focusing on individual systems. But at our company, by clarifying the whole vision at first, we are clearly defining the direction of the approach to AI so that the entire effort can proceed efficiently. The viewpoint to be clarified in the overall vision are as follows.

1: Clarify the role AI will play in manufacturing innovation in the manufacturing industry (chapter 3.1)
2: Clarify the plan for advancing AI technology and expanding the use of AI (roadmap) in the manufacturing industry (chapter 3.2)

3.1. Manufacturing Innovation Model

To clarify the role of AI in the innovation of the manufacturing industry, we defined a manufacturing innovation model (Figure 2).

In Figure 2, the targets of manufacturing innovation are divided into two levels. Level 1 is the dramatic improvement of QCD in the manufacturing process, and Level 2 is the value creation of the product itself, which is described on the horizontal axis of the figure.

Dramatic improvement in QCD is possible when the components of Level 1 have collaborated. A typical example of collaboration is the use of AI for “improvement”. In the manufacturing industry, “improvement” is hugely beneficial and has produced many effects. On the other hand, in human-centered “improvement”, the amount of information is small in the range that people can handle, making the speed of “improvement” limited due to humans’ movement and investigation. Therefore, people become the bottleneck, and it is impossible to handle a large amount of information or run “improvement” processes at high speed. In contrast, digitizing, collecting, analyzing, predicting, and controlling information using AI makes it possible to handle a large amount of information and perform “improvement” processes at high speed, thereby significantly improving QCD.

However, with the current AI technology, when comparing the “improvement” made by humans and those made by AI, “improvement” made by humans is often superior in terms of point of view. Therefore, to achieve dramatic “improvement”, it is essential to proceed with both human “improvement” and AI “improvement”. This “improvement” is defined as “super improvement” in contrast to conventional “improvement” (Figure 3).

The value creation of Level 2 in Figure 2 is to discover new concepts different from the conventional ones and therefore increase the value of the product or business model itself. 2A is to add new value to products by incorporating AI into them. 2B is to create new business models and dramatically increase their value by applying AI to the service or the businesses themselves that using the products. 2C builds a platform where many

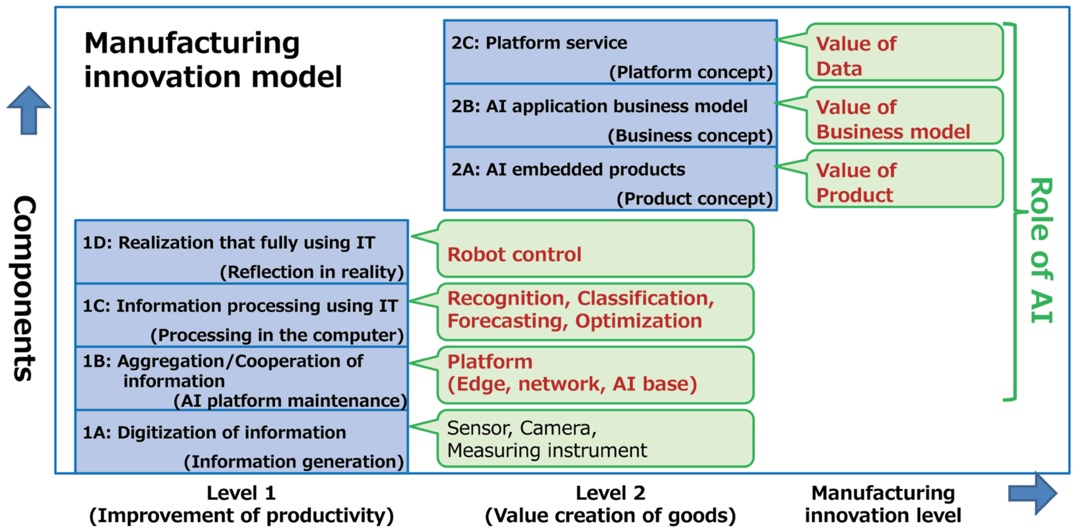


Fig.2. Manufacturing innovation model.

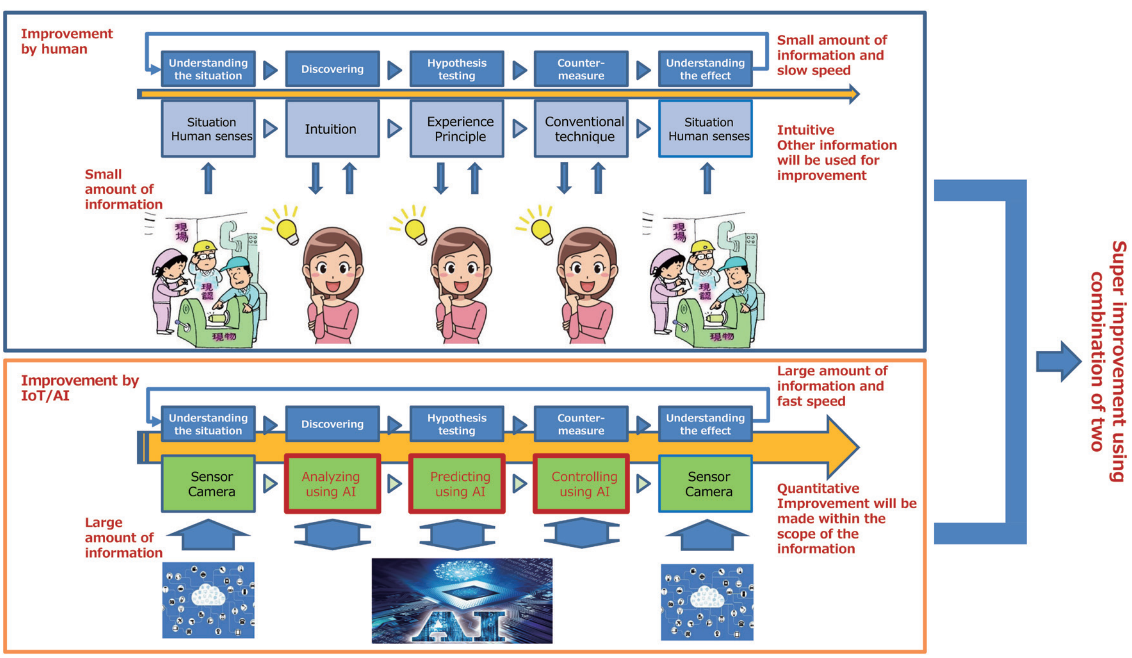


Fig.3. The improvement by human and AI.

stakeholders can coexist in one business model, or multiple business models can coexist.

In general, the role of AI does not include 1B: information aggregation/coordination, but we are promoting 1B under the leading of AI, which we call "AI platform". The "AI platform" will be explained in chapter 4.5.1.

3.2. AI System Development Roadmap

The evolution of AI is rapid, and there is a significant difference in technical difficulty. It is essential to decide the target of efforts according to our company's ability and raise the technical level. For this reason, we have created an AI system development roadmap to clarify how to advance AI technology and expand the area of AI utilization (Figure 4).

In Figure 4, the horizontal axis is the information handled by AI, and the vertical axis is the key technology used to handle the information. In addition, the areas of use that can be implemented by information and key technology are described inside each graph bar. The numbers indicate the start/scheduled time of operation of the AI system.

The information handled by AI on the horizontal axis is classified into four categories: "images", "numbers/symbols", "motion/control", and "language/concepts". Images are classified into three categories: "single still image", "single video", and "multiple images". Numbers/symbols are classified into three categories: "graphing", "time series", and "composite". These are listed in order of ease to implement in our company.

The vertical axis is the key AI technology for handling information on the horizontal axis. Specifically, Convolutional Neural Network (CNN) for single still images, high-speed technology such as distillation for single videos, object detection technology for multiple images, recurrent for time series, multimodal for composite, reinforcement learning and Simulated to Real (Sim2Real) for motion/control, and Transformer for language/concepts.

In the area of use in the graph bar, visual inspection, time series prediction, "improvement", and autonomous control are considered critical areas.

4. Actual Examples of AI Applications and Technologies

4.1. List of Projects

Based on the AI roadmap, we are gradually advancing the understanding and development of our key technologies. The projects in progress as of January 2021 (Figure 5) are shown on the same horizontal axis as in Figure 4.

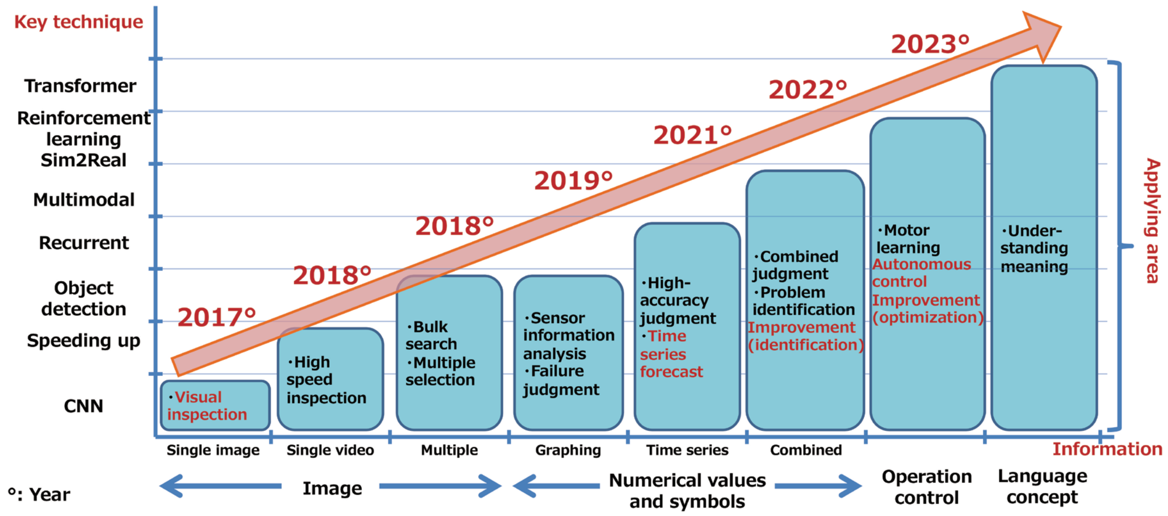


Fig.4. AI system development roadmap.

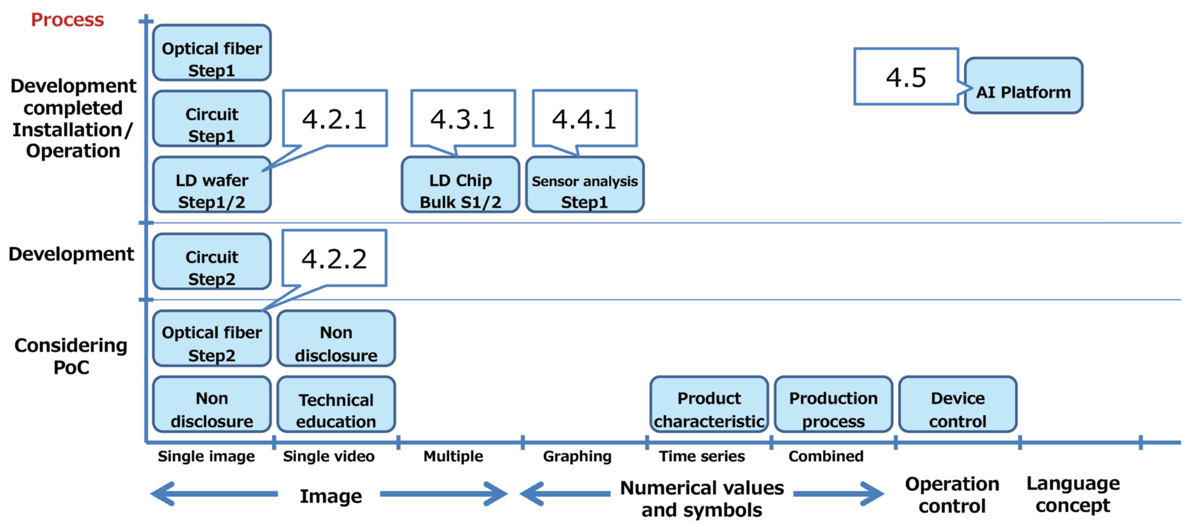


Fig.5. List of projects.

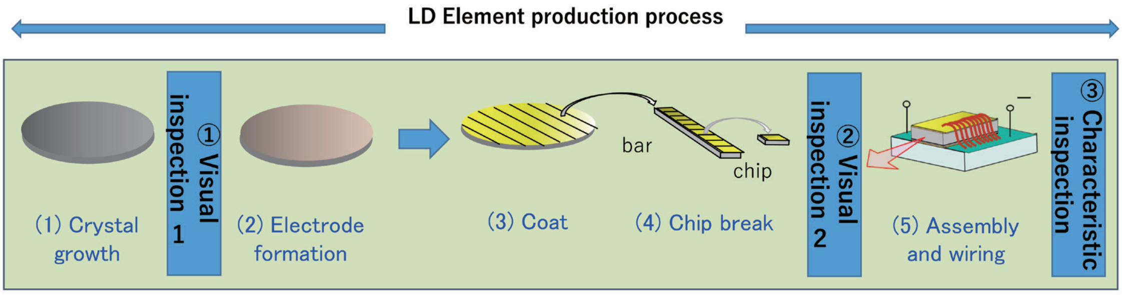


Fig.6. The production process of laser diode element.

The vertical axis shows the status (process) of the projects. Projects with chapter numbers in the figure are explained in the subsequent chapters.

4.2. Single Still Image

A "single still image" is the information that contains only one non-moving subject in the image. The key technology here is CNN, used to determine what the subject is classified as, i.e., image recognition. In the manufacturing industry, this technology is mainly used for visual inspection using images.

4.2.1. Laser Diode Wafer Visual Inspection

This chapter describes the laser diode wafer visual inspection system introduced in the high-power production process of laser diode elements. This system has been developed and operated jointly with OPTOENERGY Inc.

4.2.1.1. Overview of Operations Challenges until now

There are three inspections in the production process (Figure 6). The system in this chapter automates "(1) Visual inspection 1" by using AI. And chapter 4.3.1 will explain "(2) Visual inspection 2".

Until this system was applied, an engineer familiar with the element's structure conducted the visual inspection in a cleanroom. Since good/defective products are judged considering long-term reliability, the judgment is difficult because it depends on the combination of location, shape, size, and color (Figure 7).

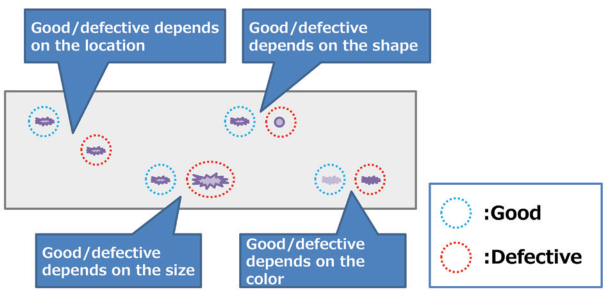


Fig.7. Rule for quality judgment.

4.2.1.2. System Overview

The hardware configuration diagram of this system is shown in Figure 8. The components of the system include an imaging device and an edge AI, which are set up in a cleanroom. The edge AI is connected to the Factory Automation Network (FA network), connected to the Internet Protocol Network (IP network) via a GW, and then to the AI platform at the AI center. The edge AI also controls the imaging device and runs on Windows OS (Table 1).

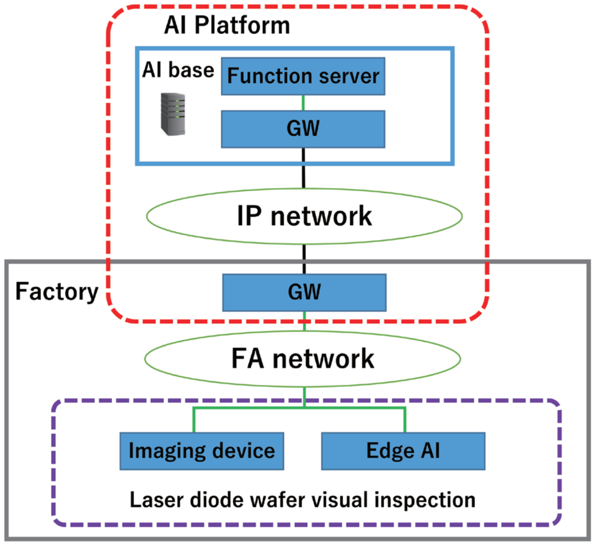


Fig.8. Hardware configuration of the system.

Table 1. Software environment of edge AI devices.

Section	Software
OS	Windows
Deep learning framework	Keras, TensorFlow
GPU platform	CUDA
Programming language	Python
Network model	EfficientNet

4.2.1.3. Status of AI Application

The system operation flow is shown in Figure 9. (1) After setting the wafer on the microscope, the system automatically moves the stage and takes multiple photographs. (2) Combine the image of the entire wafer from multiple photographs. (3) Extract the image of each chip from the combined wafer image. (4) Classify the extracted chip images using AI; at this time, to improve the manufacturing process, both good and defective products are classified into several classes. (5) Display the inspection results in a format that is easy for people to understand.

Examples of good and defective products are shown in Figure 10. The upper row is an example of a good product,

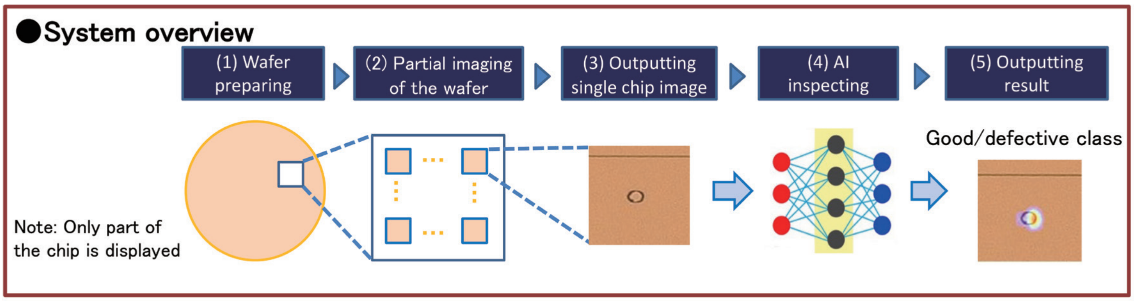


Fig.9. Operation overview.

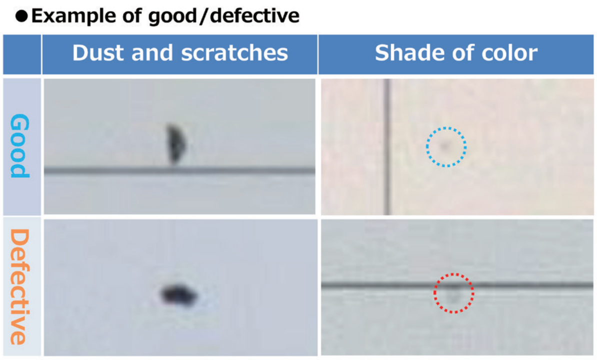


Fig.10. Example of judgment.

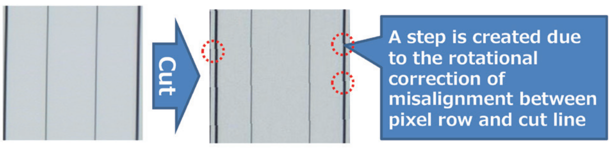


Fig.11. Misalignment problem when cutting out images.

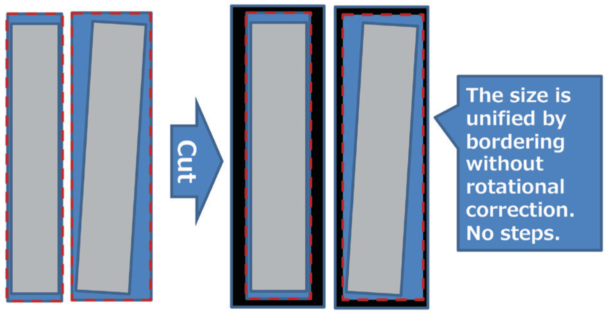


Fig.12. Picture framing method.

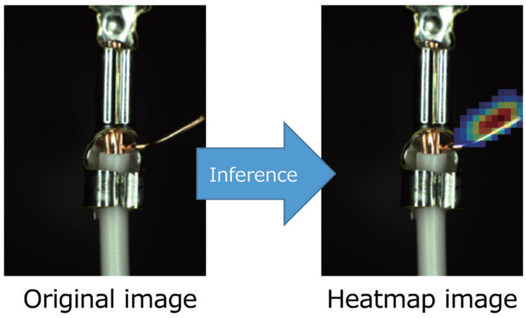


Fig.13. Visualizing the reason of inferences.

and the lower row is a defective one, with minor differences.

4.2.1.4. Implemented Technologies

We will explain three technologies implemented in this system: the picture framing method, the heatmap, and the learning progress visualization.

The picture framing method is a technique used to cut out a chip image from an entire wafer image. It is difficult to make the camera's pixel rows and the chip's edge perfectly parallel when photographing wafers. Therefore, when chips are cut out along the pixel rows, they are cut out diagonally. Furthermore, the cutout size varies depending on the degree of skewness. When analyzing images with CNN, it is necessary to unify the image size. There is a method of enlarging/compressing the image to a uniform size after cut out, but we cannot apply it to this project because detailed variations also change.

In such a case, a method is used to rotate the object to be cut out so that it is parallel to the pixel rows before cutting out. At this time, a step is generated by the rotation (Figure 11), so smoothing is often performed to make the step not noticeable.

On the other hand, when the judgment is affected by minor differences in pixel units, as in this case, accuracy will decrease when the smoothing process blunts the changes. To solve this problem, we decided to cut out the object in a unified range more extensive than the object and input it to the AI without rotating. Furthermore, when cutting out in a large size, surrounding objects may enter the cutting range, so they must be detected and filled with a specific color. This technique is called the picture framing method (Figure 12).

Heatmap is a technology to visualize which places AI focuses on when classifying. With the current technology, it isn't easy to logically show the reason for judgment. On the other hand, convincing the user of the AI judgment is challenging without showing its reason. To solve this problem, we decided to display the parts of the image that AI pays attention to when classifying them in different colors according to the level of attention. Since the laser diode chip image contains information that we must not disclose due to manufacturing reasons, we will use a crimped terminal as an example (Figure 13).

In the image on the left of Figure 13, the part where the line jumps out is the abnormal point, and humans make a judgment by focusing on this part. The image on the right shows the focus of AI using a heatmap. The left and right

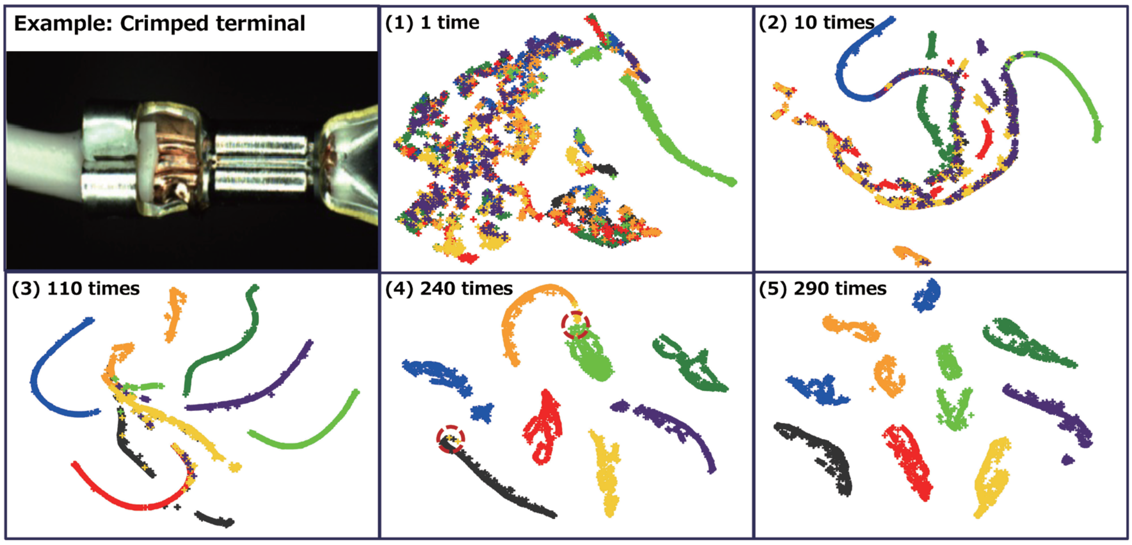


Fig.14. Visualizing the learning progress (1).

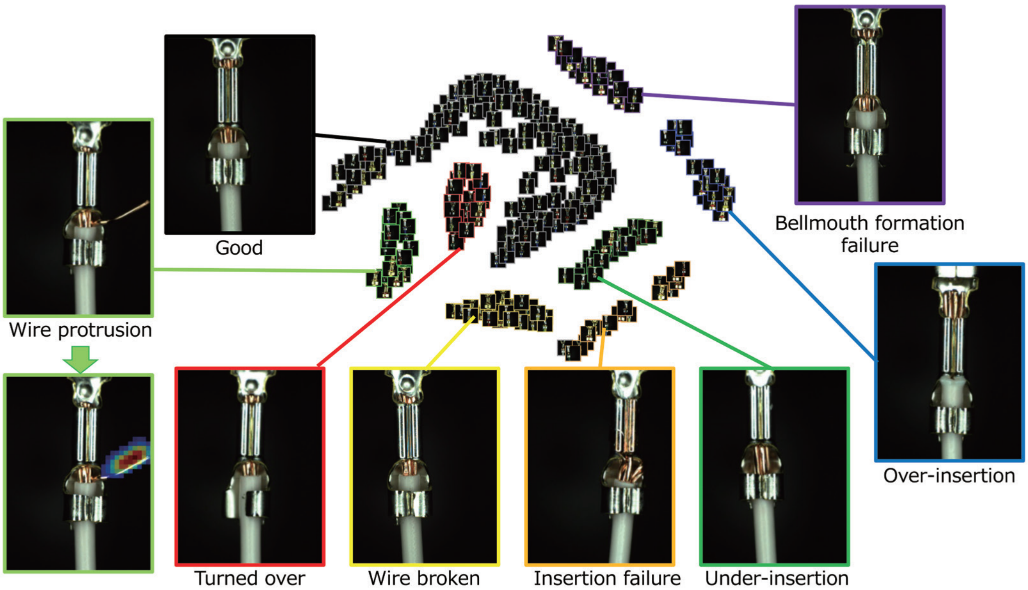


Fig.15. Visualizing the learning progress (2).

images show that humans and AI focus on the same part, and humans can visually recognize the reason for the AI's judgment.

Learning progress visualization is a visualization of the learning progress of AI using dimensional compression. Even if the learning status of AI is shared with stakeholders in the form of numbers and graphs, it is difficult to imagine the learning status, and the anxiety about AI accuracy still cannot be resolved. This difficulty is one of the reasons why it is challenging to introduce AI systems. To solve this problem, we introduced the visualization of the learning progress using dimensional compression. We will illustrate this technique (Figure 14) in the same case study as Figure 13.

Figure 14 shows the learning progress in order of the increasing number of epochs. In (4), the results are still mixed, but in (5), they are entirely separated. By sharing this diagram with the stakeholders, it is easier for them to understand the learning situation.

In addition, Figure 15 is a mapping of images to the learning progress. We can check images that are prone to mistakes and images that have not been trained well, which can improve the accuracy of AI.

We achieved 99.75% AI accuracy using the above techniques compared to human accuracy (95%) (Table 2).

Table 2. Accuracy of AI judgment.

Test: Over $4x10^4$ chips Accuracy:99.75%		A I	
		Good	Defective
Correct answer	Good	96.46%	0.12%
	Defective	0.12%	3.29%

4.2.2. Optical Fiber Visual Inspection System

This chapter describes a case study of the application of AI to recoated fiber inspection. However, the system configuration is not disclosed and will be omitted.

4.2.2.1. The Issue

When inspecting an optical fiber in detail, the inspection is performed visually. The upper part of Figure 16 is an example of a good product, and the middle/lower parts are examples of defective products.

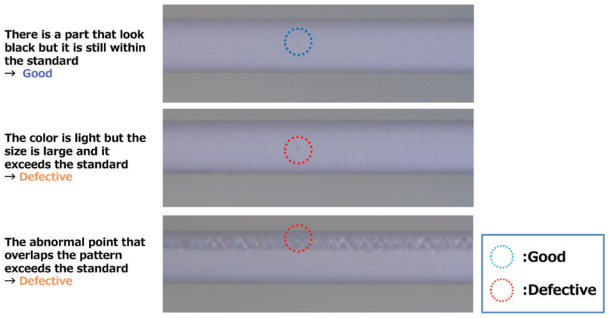


Fig.16. Judgment example of visual inspection of optical fiber.

In optical fibers, the same part may look good or defective depending on different conditions such as the position of the light source, camera angle, type of abnormal point, and position of the abnormal point (Figure 17). (1) is an example of a good product, but only the middle image looks defective. (2) is an example in which only the middle image looks good, although it is a defective product.

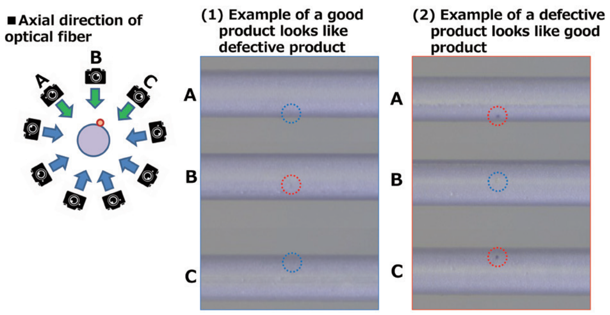


Fig.17. How the abnormal point look under different conditions.

4.2.2.2. Implemented Technologies

To solve the problem in Figure 17, we have developed a technology that combines multiple images taken under different photographing conditions into a single image and makes a judgment using AI (multi-viewpoint analysis). Figure 18 shows an example of this technology in action. By combining images taken from multiple viewpoints, we can determine whether a product is good or defective with high accuracy (Table 3).

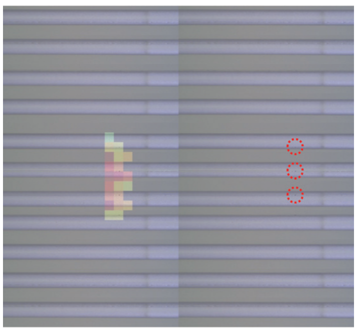


Fig.18. Judgment example of multi-viewpoint analysis.

Table 3. Accuracy of AI judgment.

Accuracy of whole circumference 98.75% (3147 cases)		AI	
		Good	Defective
Correct answer	Good	27.50%	0.0%
	Defective	1.25%	71.25%

4.3. Multiple

"Multiple" is the information that combines multiple subjects in an image, regardless of whether they are moving or not. The key technology here is object detection, which can identify the subjects. In the manufacturing industry, it counts the number of subjects and judges the quality of individual subjects.

4.3.1. Laser Diode Chip's Batch Inspection System

In this chapter, we describe the laser diode chip's batch inspection system. We automated "(2) visual inspection 2" as described in Figure 6 by using AI. The system configuration and operation flow are almost identical to those described in Section 4.2.1, so they are omitted. This system is also developed and operated in collaboration with OPTOENERGY Inc.

Examples of good and defective chips are shown in Figure 19. The upper row is an example of a good chip, and the lower row is a defective chip. Figure 20 shows the inspection results. If the detected chip is good, it is displayed in green, and if it is a defective chip, it is displayed with a red frame. In a chop break error, various chip shapes and fragments are generated, making it difficult to detect objects reliably using conventional techniques based on template matching.

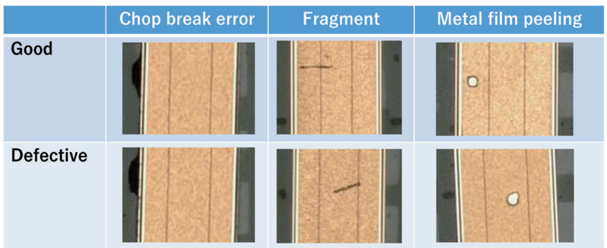
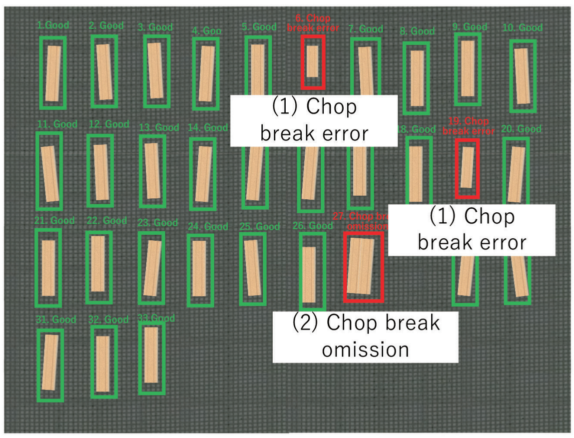


Fig.19. Example of good and defective products.



Note: The shape of chip is for illustration

Fig.20. Display of inspection results.

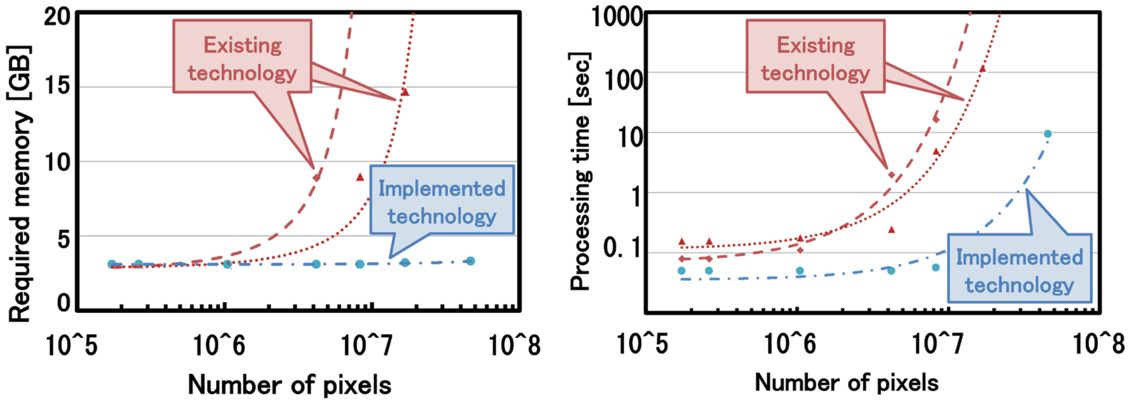


Fig.21. Resources and processing time of object detection.

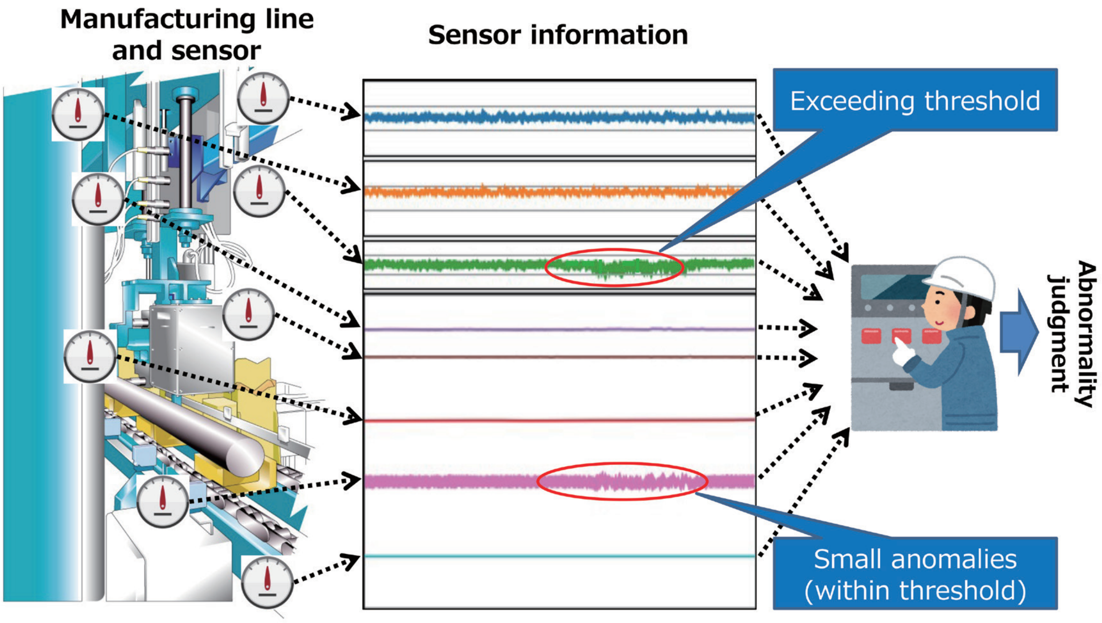


Fig.22. Sensor data analysis.

4.3.1.1. Implemented Technologies

As the image size increases, the amount of memory and processing time required for object detection using AI increases, resulting in higher system costs. To resolve this problem, we separate object detection and class judgment and then perform AI judgment with an image size suitable for each detected chip, reducing computer resources and improving judgment accuracy compared to object detection without this technique. In Figure 21, (1): the inspection image is taken with large size of about 50 million pixels because the abnormal points in the inspection object are small. (2): Since object detection with the image size of (1) requires a lot of computer resources, the image is compressed to about 90,000 pixels, and (3): perform only object detection. (4) Return the detected inspection object to the original image size and perform class judgment. Using this technique, we achieved high judgment accuracy while reducing the computer resources (Table 4).

Table 4. Visual inspection 2: accuracy.

Object detection rate	100.0 %
Good/defective accuracy rate	99.6 %

4.4. Graphing

"Graphing" is the information obtained by converting numerical values/characters/symbols into graphic positions/shapes/colors. The key technology is CNN. In the manufacturing industry, we can use it to determine whether a product is in good or defective condition based on sensor output information and various parameters.

4.4.1. Sensor Data Analysis System

This chapter describes the sensor data analysis system.

4.4.1.1. Overview of Operations and Challenges until now

Many sensors are installed at the manufacturing site. There is a task to aggregate/accumulate these data and make a time series graph to determine whether or not there is an abnormality. Furthermore, judgments are made not only based on the threshold value but also on the subtle variations in the graph (Figure 22). To achieve this with conventional technology, we need to define small anomaly and create detection algorithms. Still, it is difficult to create all the algorithms because of the variety of anomaly. Although we have relied on skilled workers, it is increasingly difficult to secure human resources due to the declining birthrate and aging population. To solve these problems, we have developed this system.

4.4.1.2. System Configuration

The software configuration is shown in Table 5. The hardware configuration is omitted.

Table 5. Software environment of edge AI devices.

Section	Software
OS	Ubuntu
Deep learning framework	Keras, TensorFlow
GPU platform	CUDA
Programming language	Python
Model	MobileNet v2 base

4.4.1.3. Status of AI Application

The sensor often operates continuously for a long time, which results in a long graph. It also increases the possibility that there are multiple abnormal points in the graph. Therefore, to apply AI, we decided to divide the sensor data at regular intervals into multiple segmented graph images and tag, learn, infer each segmented image. However, we overlap the images by a certain width to prevent the accuracy from being degraded because the abnormal point is split into two separate images. The fixed interval was determined considering that the divided image should not contain multiple abnormal points, and a single abnormal point should not straddle multiple divided graph images as much as possible (Figure 23).

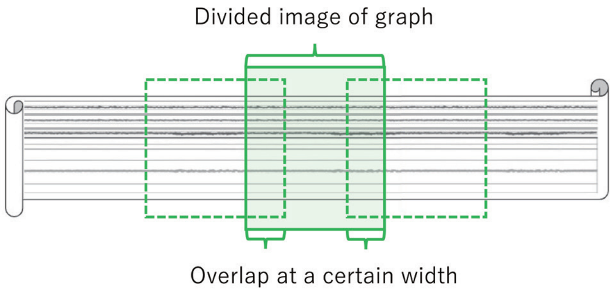


Fig.23. How to divide the graph.

4.4.1.4. Implemented Technologies

In general, it is difficult for CNN to detect the subject position precisely. In this system, to perform standard value judgment precisely, we developed a technique to improve the judgment accuracy by drawing a guideline in the graph to indicate the threshold value and having AI learn the relationship between the guideline and the graph (Figure 24). The judgment accuracy is shown in Table 6.

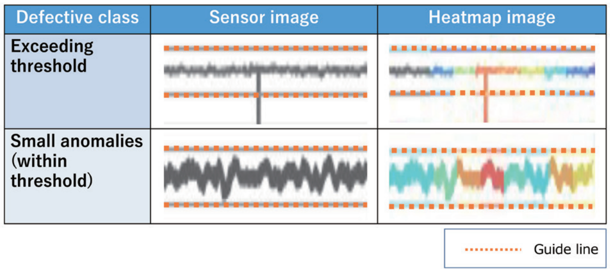


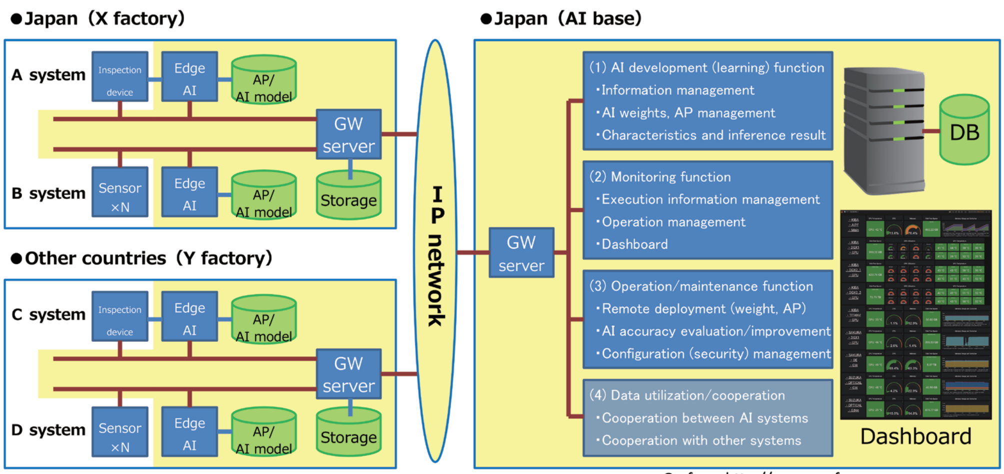
Fig.24. Guideline and judgment result.

Table 6. Accuracy of AI judgment.

		A I	
		Good	Defective
Correct answer	Good	31.2%	0.2%
	Defective	0.5%	68.1%

4.5. AI Platform

The AI platform is being developed and operated to increase the value of AI systems by optimizing their development, monitoring, operation, and maintenance. As shown in the configuration diagram (Figure 25), the AI platform consists of the AI systems installed in factories in Japan and overseas, the functional servers installed in AI bases, and the networks connecting them. The functional servers consist of four functional groups: (1) AI development (learning), (2) monitoring, (3) operation and maintenance, and (4) data utilization/linkage. The AI platform is implemented on deep learning supercomputers (NVIDIA DGX-1 and DGX-2). The AI platform allows remote control of AI systems at factories from the AI base, and by assembling AI engineers at the AI base, it is possible to reduce costs and respond quickly.



Grafana, http://www.grafana.com

Fig.25. Configuration diagram of AI platform.

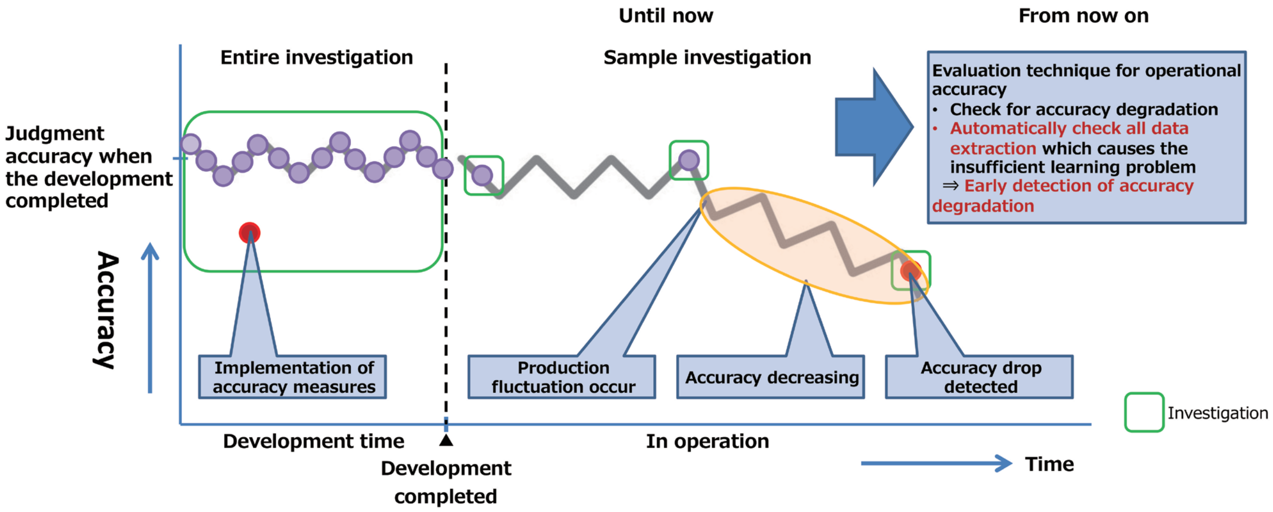


Fig.26. AI accuracy evaluation during operation.

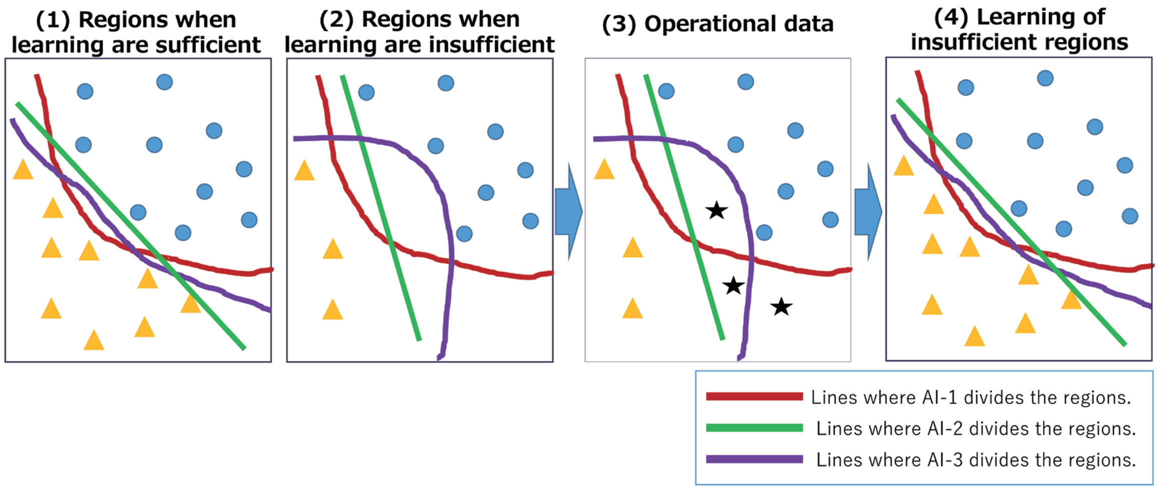


Fig.27. Difference in accuracy due to different weights.

4.5.1. AI Accuracy Evaluation/Improvement during Operation

To improve the accuracy of AI, we would like to prepare all possible information at the time of development, but in reality, this is extremely difficult. For example, after the start of the operation, changes in environmental conditions such as temperature/humidity, fluctuations in manufacturing conditions (minor changes within standard values), and changes in 4M (Man, Machine, Material, Method) may occur. As a result, there is a concern that judgment accuracy may decrease due to insufficient learning or unlearning. In response to this concern, it is common to check the accuracy of AI by extracting samples after the start of operation and comparing the judgments between AI and humans. However, increasing the number of samples will increase the amount of human work and decrease the effect of the AI system. On the other hand, reducing the number of samples increases the risk of delay in detecting a decrease in AI accuracy (Figure 26).

To solve this problem, we developed a technique for evaluating AI accuracy during operation. This technique takes advantage of the fact that even if the deep learning network structure and training data are the same, the weights generated by random factors in the training process differ slightly, resulting in differences in judgment accuracy and judgment strengths and weaknesses. In particular, this difference tends to occur at class boundaries where there is little training data.

This technique is illustrated in Figure 27. The three lines divide the yellow triangle and blue circle regions by AI with different weights. (1) is an example of sufficient learning: the three lines divide the regions almost the same way, and there is no difference in judgment. (2) is an example where each weight has a different boundary, resulting in a difference in judgment. Therefore, the judgment result of the black star mark in (3) differs depending on the weight.

The data whose judgment result differs depending on the weights become candidates for insufficient training. In (4), the judgment difference is eliminated by carefully examining the candidate data and performing additional learning. Furthermore, since the candidate data generation rate correlates with the percentage of insufficient training of AI, i.e., the error rate, we can also use it to estimate AI accuracy.

This technology is implemented and operated in the AI system developed at our company. We calculated the true AI accuracy by comparing the judgments between humans

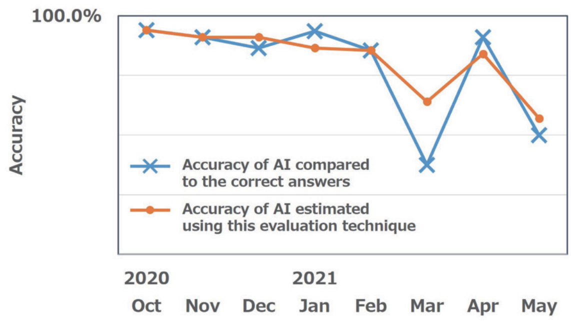


Fig.28. Result of accuracy evaluation.

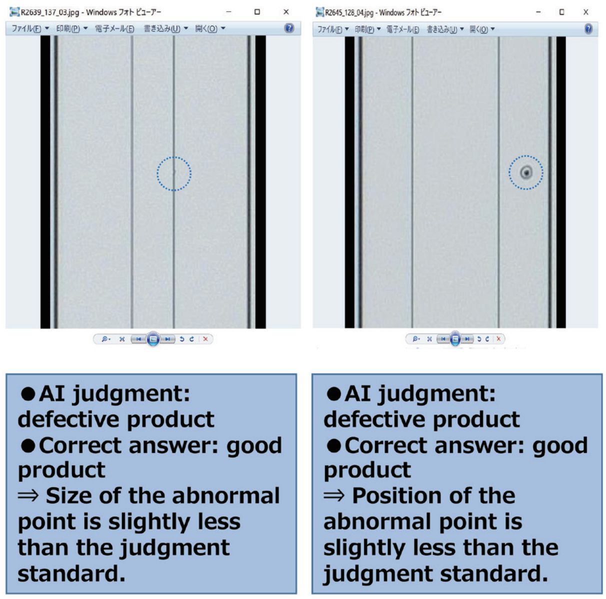


Fig.29. Examples of images extracted by this technique.

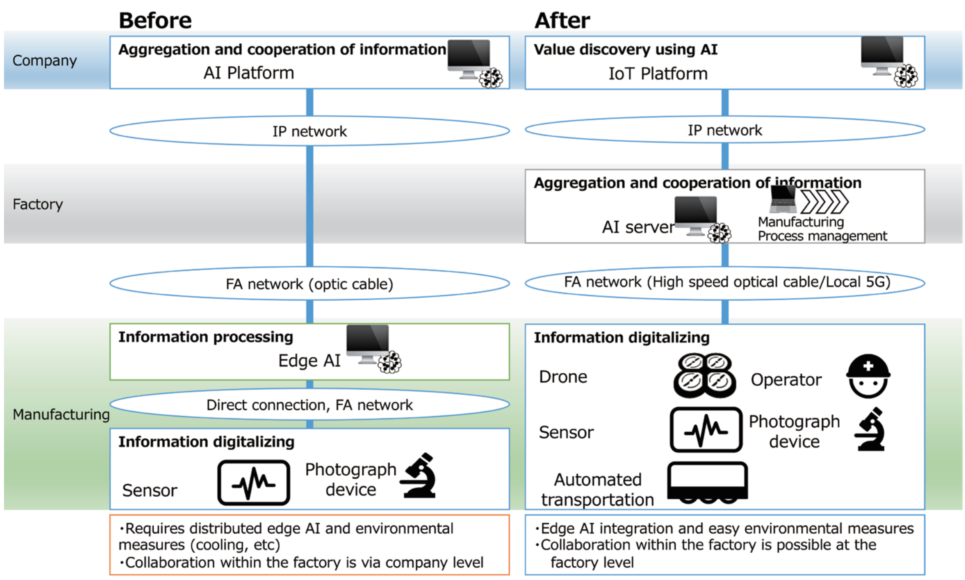


Fig.30. Changes in AI system architecture.

and AI of 10% sampling results from the operation period. At the same time, we applied this technology to the sample target, estimated the AI accuracy, and obtained a high correlation (correlation coefficient of 0.93) (Figure 28).

In addition, when candidates with insufficient learning were extracted using this technique for all examinations during the same period, we confirmed that 70% of the candidates were insufficient. The extracted images are shown in Figure 29.

5. Changes in the AI System Architecture according to the ICT Evolution

In parallel with the evolution of AI technology, ICT is also evolving. In wireless communications, 5G has begun, and the speed, low latency, and the number of simultaneous connections have improved significantly compared to 4G and Wi-Fi. In addition, edge AI is becoming more powerful, equipped with GPUs to process complex AI models at high speed. These ICT advancements are changing the AI system architecture.

The left side of Figure 30 shows the current AI system architecture. The problem is that the edge AI is distributed in the manufacturing layer, making it difficult to manage, maintain, and ensure security. Furthermore, to install a high-performance edge AI which generates a large amount of heat in a factory where environmental conditions (temperature/humidity/voltage fluctuations/vibration/dust, etc.) are a challenge, we install the edge AI in an air-conditioned box to protect it from high temperature/high humidity/vibration/dust, and a Uninterruptible Power Supply (UPS) must be equipped to supply stable voltage and current. In addition, AI models are becoming more complex every year, and edge AI will become more powerful, and the specifications are expected to be further improved in the future.

Furthermore, as for the layers where computers are installed, it becomes difficult to isolate them in case of abnormalities if the number of installed layers is too large, so currently, computers are installed in the manufacturing layer and the company-wide layer. However, with this architecture, it is necessary to perform the collaborative processing within the factory at a company-wide layer, leading to wasteful communication and reduced reliability. For example, a failure of the IP network may cause the interworking process in the factory to stop.

The reason for these problems is that to make the communication between the information digitizer and the edge AI with high speed, low latency, and it is necessary to use high reliability, direct connection, or FA network (optical), and it isn't easy to separate the information digitizer from the edge AI.

The evolution of ICT will allow the introduction of low latency industrial Ethernet, high-speed optical fiber, and 5G into FA networks, which will solve the current problems. The right side of Figure 30 shows the direction of AI system architecture in the future. With the advancement of communication technology, edge AI can be integrated into the factory servers, making it easier to manage and process in a server room with stable environmental conditions. The information digitizer and the factory server are connected by a high-speed optical fiber or local 5G, depending on the frequency of changes in the production line configuration and equipment transferring. In some cases, both high-speed optical fiber and 5G are used to ensure communication reliability. Furthermore, computers will be installed at the factory layer and the company-wide layer, facilitating the execution of coordinated processing within the factory.

We will continue to build flexible and scalable systems for the future evolution of ICT and AI system architecture, such as loosely coupled edge AI and information digitizers.

6. Conclusion

In this paper, we explained the application of AI to manufacturing processes at our company and clarified that AI is in the practical stage in the manufacturing industry. We also discussed the evolution of the architecture of 5G and other technologies. In the future, we will follow our roadmap of AI system development and further accelerate efforts to expand the information handled by AI and the scope of its application in our business.

7. References

1) Masahiro Kashiwagi, et. al.: "Approach to practical application of Deep Learning in manufacturer's production line", GPU Technology Conference, 2017.
2) Kiminori Kurosawa: "AI (Deep Learning) efforts at Fujikura and its application to the visual inspection of high power semiconductor laser wafer", GTC-Japan, 2017.
3) Hiroyuki Kusaka, et. al.: "Practical Application of Deep Learning in Smart Factory: Visual Inspection System of Semiconductor Laser", GTC-San Jose, 2018.
4) Hiroyuki Kusaka, et. al.: "High-accurate visual inspection for semiconductor lasers using Convolutional Neural Network trained with original small dataset and additional synthesized images", IOT Enabling Sensing/Network/AI and Photonics Conference, 2018.
5) Shinichi Nakatori, et. al.: "Automation of Inspection Process by AI (Deep learning)", IoT-SNAP, 2019.
6) Kazuichi Tada: "Collaboration play between ordinary AI and excellent AI: Automating semiconductor wafer visual inspection", Nikkei XTREND, 2018.
7) Japan Deep Learning Association: Textbook of Deep Learning Application - Practical Edition, Nikkei BP, 2018.
8) Japan Deep Learning Association: Textbook of Deep Learning Application - Practical Edition, Nikkei BP, page 60, 2019.