2025 IMEKO TC-6 International Conference on
Metrology and Digital Transformation - **M4DConf 2025**
Benevento, Italy, September 3-5, 2025

Development and Application Study of a Calibration Certificate Anomaly Detection System

Dong-Hun Ryu[1], Chae-Wook Lim[1]

[1] *Korea Testing Laboratory, 723 Haean-ro Sangnok-gu Ansan-Si, Gyeonggi-do, 15588, South Korea, dhryu@ktl.re.kr, hicor1@ktl.re.kr*

This study focuses on the development of a system to effectively extract and utilize data from approximately 3,000 types of unstructured calibration certificates (e.g., paper, PDFs) widely used in the calibration industry. To process document-based data that are difficult for machines to recognize, the Korea Testing Laboratory (KTL) has developed a calibration certificate anomaly detection system called *K-Argos*, which applies object detection AI technology based on Convolutional Neural Networks (CNNs).
The core function of the system is to extract data from unstructured documents, convert them into structured data, and standardize and store the results.
Using the converted structured data, *K-Argos* provides statistical data management features such as performance analysis of calibrated equipment and anomaly detection in calibration data. Additionally, the system can be used to correct inaccurate identification information (e.g., manufacturer, model, serial number) and to analyze errors within the calibration data.
In the future, the *K-Argos* system will be expanded to include an online technical supervisor service that automatically determines anomalies in data during the certificate approval stage and provides relevant information to the approver (technical supervisor). This is expected to significantly improve the efficiency and reliability of calibration tasks and contribute to the digital transformation of the calibration industry through the implementation of Digital Calibration Certificates (DCCs)

I. INTRODUCTION

Korea Testing Laboratory (KTL) is a public institute under the ministry of trade, industry and energy of South Korea and in charge of test and certification for industries. Among the nine divisions, the industrial standards division is performing calibration of measuring instruments, legal metrology and reference materials. The division issues approximately 120k calibration certificates annually and recently developed a tool to digitalize and utilize the calibration data to enhance the quality of calibration service.

The main works can be summarized as follows.

1. DB accumulation using OCR
2. Grouping similar results with identifier (manufacturer, model, S/N) correction
3. Grouping similar results with nominal value estimation
4. Analysis using the accumulated results

Two main technologies are OCR and grouping.

Currently, it is planned to develop the following services further to improve the internal efficiency and customer experience.

1. Implementation of the developed anomaly detection system in the certificate approval process
2. The client service to digitalize paper certificate
3. Related terms search feature

II. DB ACCUMULATION USING OCR

In case of KTL, the first page of calibration certificates contains the information required by ISO 17025 [1] such as title, the name and address of the client, the location of the lab, etc. It is relatively easy to make a structured database because the format is similar in all fields. But it is difficult to accumulate the data and manage it as database from the second page because each tester uses his/her own format and software (word, excel, etc.) to express the results and test method.

To develop a service with calibration results, it is preferred to have enough data. But there was a lack of human resources to make a database of legacy certificates already issued, manually. As an alternative, we developed an OCR tool for data extraction from pdf certificates of the past, which utilizes the object detection AI based on convolutional neural network. 120k certificates can be categorized to 3000 formats and we trained an AI model for most common 100 formats and made a database of 100k+ certificates. The figure 1

shows that various formats are used to show the results. While the first page always uses similar format, the second page contains different number of data in different tabular forms. This causes difficulties in digitalizing the certificate using OCR, because the number and location of target data changes. Coding each case is also very inefficient because too many cases exist. The figure 2 shows the developed OCR tool.

The procedure of OCR tool can be divided into two steps. The first step is a table detection, and we used the Cascade TabNet model.[2] From the 100 formats, 1000 cases were generated by image distortion and used for learning. The table detection was very successful, and the image was converted to low resolution, black and white image to lower the cost of second step using LLM. The second step is the table structure recognition, and we used GEMMA 3[3] and EXAONE[4]. GEMMA 3 is open-source AI model developed by GOOGLE and EXAONE is open-source AI model developed by LG AI Research. EXAONE has competency in understanding Korean language and GEMMA 3 and EXAONE are used via LangChain.[5] The data extracted by GEMMA 3 are sent to EXAONE and recognized as nominal value, reference, indication, error and uncertainty.

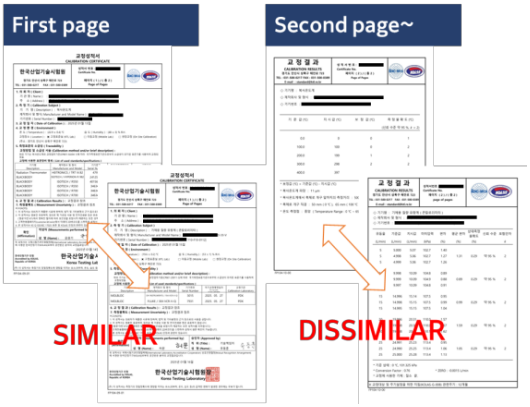

Fig. 1. Example of calibration certificates that shows the format difference of results page

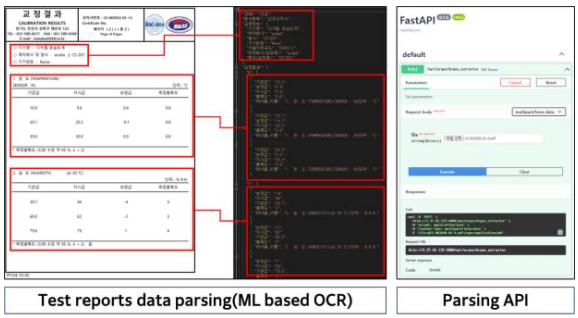

Fig. 2. Developed Machine Learning based OCR and API

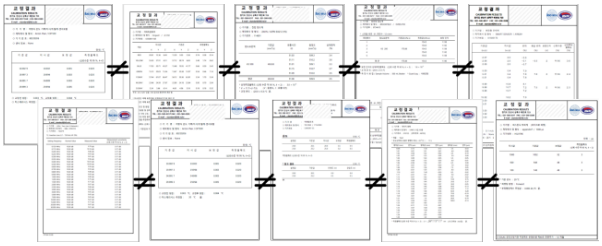

Fig. 3. Example of calibration certificates that shows the various formats depending on measurement field

Figure 3 shows that a very wide range of formats are used depending on the measurement field, and different labels are used quite often. For example, “the conventional mass” is the label used for the “reference value” in mass weight certificate and sometimes table has different structure and merged cells. We focused on realizing database establishment of past certificates using OCR and testing possible service with this DB. The success rate is difficult to know because we could not test all 3000 formats. As a first step of applying OCR and DB establishment, we focused on the most used formats, and it was successful for 100 formats, and we could make a database of 100k+ certificates. This required a lot of prompt engineering to make it understand that “conventional mass” can also mean “reference value” sometimes. We are planning to expand the applicable formats in future, and some formats are very difficult to recognize and discussion with field experts for standardization is needed.

III. IDENTIFIER CORRECTION

To analyze the results, it is essential to identify and group the results. The manufacturer, model and serial number are very good identifiers of measurement instruments, but we encountered some problems with grouping based on the forementioned identifier. There were many typos or different ways of expressing it. For example, a viscometer manufacture “BROOKFIELD” was sometimes expressed as “Brookfield”, “brook-field” and model “QL200-MK” was “QL200 MK” or “QL200MK” This small difference of capital letter, dash and space made a trouble in identifying instruments and caused errors and inefficiencies.

So, we developed similarity search system to make a correction of identifier to make it possible to find the name even when no exact matches are found. This also could be applied when the number of strings is different, or there are typos in two words. The following table shows the correction examples.

Table 1. Correction example

As is	Corrected
BROOKF**EI**LD	BROOKFIELD
TRANSFO**MM**ER	TRANSFOMER
PO**VV**ER	POWER
D**II**AGNOSTIX	DIAGNOSTIX

In developing this similarity search system, we used our database which has 70000 manufacturers and model names based on past certificates. After correcting the blank and capital letters, the data is categorized into the names and then the most used name is identified as the right manufacturer name. For example, if the most common name is “BROOKFIELD”, then “Brook-field”, “BROOKFIELD©”, “Brookfield™” will be recognized as the most common name and will not be recommended either.

The reason we mention “recommendation” is that we used this similarity search not only for grouping the calibration results data but also for recommending in self-query for customers. When customers request a query, this can recognize the misinput of manufacturer or model name, and shows the right results.

To accelerate the search, the algorithm works in sequential steps. Exact match is performed first which is the fastest, and then partial match (%like%) search comes next. It calculates cosine similarity, and we used the threshold value of 0.8. Average search time is 30 ms.

IV. NOMINAL VALUE ESTIMATION

After the grouping of the calibration certificates of the same manufacturer and model and serial number, it is required to group the values to compare the results. Reference value varies slightly, and it is sometimes difficult to make a group based on the value. Nominal value is more distinguishable and helpful when making a group of similar range results.

Additionally, it was found that each field uses different words in the certificate such as “reference value”, “conventional mass”, etc. for same meaning. For practical reasons, we decided to minimize the database category and standardized DB names. In case of 100 formats, we identified only “setting value”, “reference value”, “indication value”, “corrector or error” and “uncertainty”

V. ANOMALY DETECTION SYSTEM

We developed a calibration support tool with this data. The figure 4 is a screenshot of a system being developed by the end of 2025. From the database of calibration certificates of 5 years, annual reports for the same device or same model are sorted and analyzed statistically to detect anomalies. It shows which calibration certificate is suspected to have a typo in manufacturer, model or serial number. The right rectangle is for the case when measurement error seems to be too large when compared to the same model results. The technical manager or calibration staff can click a certificate number and check the details.

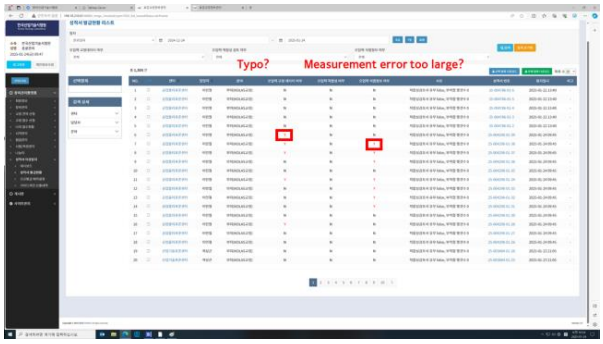

Fig. 4. Screenshot of anomaly detection system

When someone clicks a certificate number, the screen changes to the details screen, and shows the results as shown in the figure 5. This screen is an example of a torque wrench, and there were 125 results of same model in the database as shown in the figure 6. The bottom left graph shows the previous results of same serial number, which means, the same instrument. The bottom right graph shows the guideline range of the same model, and the dots in the middle show that the results are within the range.

This system is being developed for internal use and in the beginning stage. The guideline range was setup as 75 % and 25 % of previous same model results arbitrarily for now, and we plan to get feedback from the staff to find better ways in guideline set-up next year.

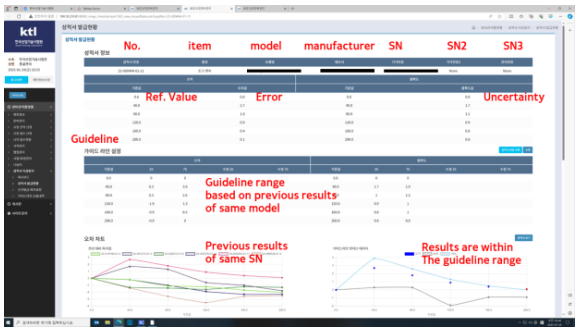

Fig. 5. Details screenshot of anomaly detection system

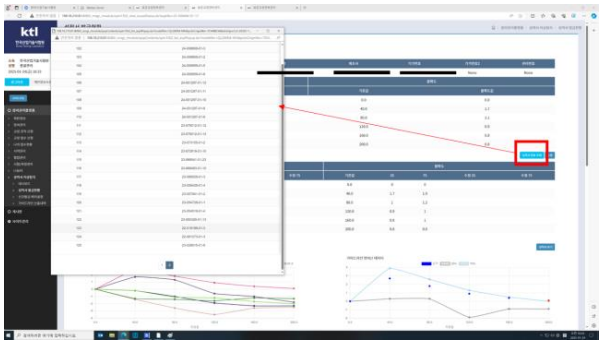

Fig. 6. The list of certificates used in the guideline setup

One more example is a micro pipette in figure 7. There is only one line in the left bottom, and it means there is no record of same pipette for the last 5 years. The bottom right graph is guideline based on the results of same model and the number of certificates used was 484. The red dot is outside the range and possibly it could be typo or test error or instrument failure. The technical manager can check it, and we think it would be a good start to give people a clue to detect anomaly. As mentioned before, the guideline range needs to be elaborated and currently manual input by each field expert is activated with default value of 25 % and 75 % value of previous results

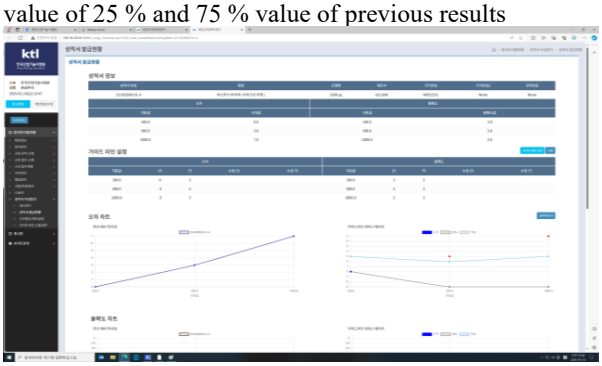

Fig. 7. Example for results outside of guideline

VI. FUTURE PLAN

A. Pre-check service

Usually, the calibration follows the (1 ~ 5) steps of the following figure. It is sometimes difficult to check all calibration certificates and every result and model and manufacturer in it because of workload. Our plan is the implementation of the anomaly detection system between step 3 and step 4 so that the calibration staff or the technical manager can easily detect the calibration certificate that seems suspicious and focus on it to check. The system is named K-argos because it is used to monitor errors in calibration certificates. Argos is a giant with 100 eyes in myth and often used for the name of monitoring system, and K is from KTL.

1) A calibration staff performs calibration
   ⇩
2) Prepares a calibration certificate.
   ⇩
3) Submit a calibration certificate for approval.
   ⇩
   + Certificate Pre-Check Service :
   detects errors in calibration data (incorrect results, equipment information, certificate forms, etc.) and notifies key review points
   ⇩
4) The technical manager reviews the certificate and makes the final approval.
   ⇩
5) The approved certificate is provided to the customer.

Fig. 8. Pre-check service scheme

B. Related terms search feature

Though some different expressions were found in the certificates (QL200-MK, QL200 MK, etc.), much wider range of expressions are used in the field. This causes difficulty when a client wants to find the right category for his/her instrument to register or request a quote. We mentioned that the similarity search was useful not only in grouping results but also in the customer query process in section III. We are going to expand the function to the related terms search feature and the idea is that we are going to make an expanded database using LLM. We let KeyLLM analyze the database and extracts related keyword pairs. Then KeyLLM expands the database including related keywords, and this is used in the search process when exact or similar match is not found. For example, “nogisu” is often used in the construction field of Korea, and the related terms search feature can recognize conceptual similarity and recommend “vernier calipers”.

C. Certificate stream service

We are reviewing the plan to offer the clint the OCR technique used in the digitalization of historical calibration certificate. With the digitalized calibration certificates inside KTL, it will be possible to provide a certificate stream service to the customer. The following figure shows the schematic diagram of the service. Currently, the KTL network is isolated for security reasons, and we made “k-tools” website for customer service. The plan is to establish APIs, and a digital calibration certificate will be provided if authorized customers log in to the k-tools and request it. The customer can digitalize his/her own calibration certificates in pdf form and store them in k-tools for

convenience. However, it is difficult to guarantee the success of xml conversion because user created OCR pdf certificates will have very wide variety. One possible plan is to provide excel conversion of pdf for convenient data handling with limited liability.

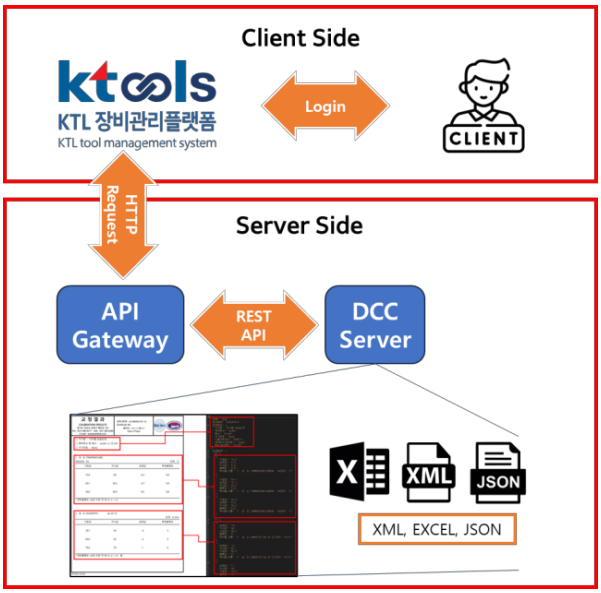

Fig. 9. Certificate stream service scheme

VII. CONCLUSION

KTL issues 120k+ calibration certificates every year and plan to digitalize certificates and make use of results. We trained AI to extract data from past pdf certificates and accumulated DB. Then, we made a correction module to identify manufacturer & model & SN. The data were sorted based on the nominal value and calibration support system was developed.

We plan to implement the calibration support system in the internal system for staff, and provide the certificate stream service and the related terms search service for external customers.

REFERENCES

[1] ISO/IEC 17025:2017 General requirements for the competence of testing and calibration laboratories
[2] https://github.com/DevashishPrasad/CascadeTabNet
[3] https://deepmind.google/models/gemma/gemma-3/
[4] https://www.lgresearch.ai/exaone
[5] https://www.langchain.com/