

Taylor & Francis
Taylor & Francis Group

Science and Technology of Advanced Materials: Methods

ISSN: 2766-0400 (Online) Journal homepage: www.tandfonline.com/journals/tstm20

Development of an automated evaluation system for the process–structure–property relationship of structural metallic materials towards autonomous exploration

Yulin Xie, Sayaka Sekida, Masaya Inakawa, Ryo Tamura, Tadashi Furuhara & Goro Miyamoto

To cite this article: Yulin Xie, Sayaka Sekida, Masaya Inakawa, Ryo Tamura, Tadashi Furuhara & Goro Miyamoto (2026) Development of an automated evaluation system for the process–structure–property relationship of structural metallic materials towards autonomous exploration, Science and Technology of Advanced Materials: Methods, 6:1, 2735657, DOI: 10.1080/27660400.2026.2735657

To link to this article: https://doi.org/10.1080/27660400.2026.2735657

© 2026 The Author(s). Published by National Institute for Materials Science in partnership with Taylor & Francis Group

View supplementary material

Published online: 22 Oct 2026.

Submit your article to this journal

Article views: 1085

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tstm20

SCIENCE AND TECHNOLOGY OF ADVANCED MATERIALS: METHODS
2026, VOL. 6, NO. 1, 2735657
https://doi.org/10.1080/27660400.2026.2735657

Taylor & Francis
by informa

OPEN ACCESS Check for updates

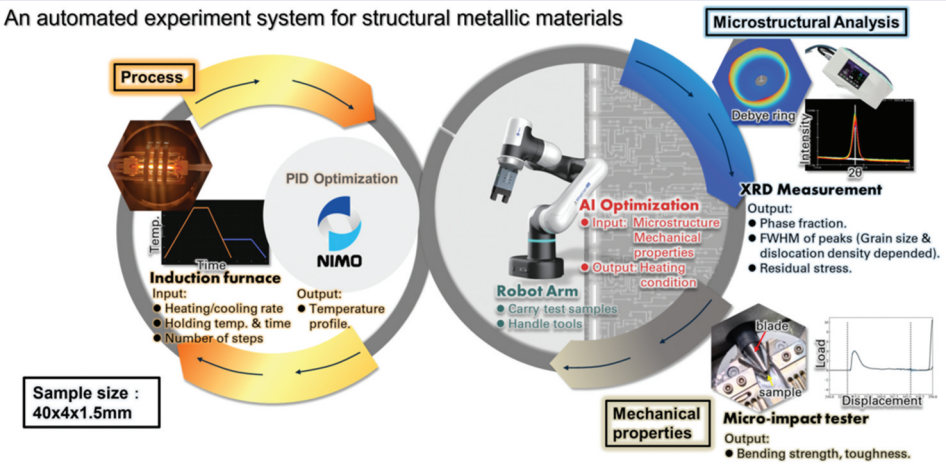
Development of an automated evaluation system for the process–structure–property relationship of structural metallic materials towards autonomous exploration

Yulin Xie[a], Sayaka Sekida[a], Masaya Inakawa[b], Ryo Tamura[c], Tadashi Furuhara[a,d] and Goro Miyamoto[a]

[a]Institute for Materials Research, Tohoku University, Sendai, Japan; [b]LaboRobo Inc., Sendai, Japan; [c]Center for Basic Research on Materials, National Institute for Materials Science, Tsukuba, Japan; [d]Research Center for Structural Materials, National Institute for Materials Science, Tsukuba, Japan

ABSTRACT

Automated autonomous experiment systems in material science are desired for exploring numerous conditions, but barely any examples for solid bulk metals are reported. In this study, an automated experimental system for structural metallic materials is designed to evaluate the process–structure–property relationship of structural metallic materials. Fully automated heat treatments, X-ray diffraction (XRD) measurements and impact testing on bulk metallic samples with a size of $40 \times 4 \times 1.5\ mm^3$ are realized by integrating a robot arm. The experimental process, excluding heat treatment, can be finished within 800 s for 1 sample. Automated high-throughput experiments on the one- and two-step tempering of a JIS SCM435H alloy are performed to demonstrate the capability and efficiency of the system, and the accuracy of the resulting property data is validated by manually measured hardness. It is also found that the diffraction peak width is positively correlated with strength, indicating the capability of the system to achieve structure–property relationships. An AI optimization tool built in the NIMS Orchestration System (NIMO) is implemented, thereby providing a basis for future closed-loop autonomous optimization.



IMPACT STATEMENT

This work present a fully automated robotic platform integrating heat treatment, X-ray diffraction, and mechanical testing for bulk structural metals, enabling AI-driven autonomous exploration of process-structure-property relationships.

ARTICLE HISTORY
Received 18 May 2026
Revised 24 August 2026
Accepted 15 September 2026

KEYWORDS
Automated experiment; structural materials; heat treatment; deformation fracture behaviour; steels; robotics

1. Introduction

In current development materials research, the scope of experimental conditions to be explored in both synthesising and processing is rapidly expanding, creating a spacious range for experimental exploration. On the other hand, due to time and economic constraints in the real world, explorations and evaluations performed by human researchers have been limited to only a considerably small fraction of these possibilities so far. Therefore, there is a growing need to realize automated, autonomous experimental systems in materials

CONTACT Yulin Xie yulin.xie.d1@tohoku.ac.jp Institute for Materials Research, Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai 980-8577, Japan

Supplemental data for this article can be accessed online at https://doi.org/10.1080/27660400.2026.2735657

© 2026 The Author(s). Published by National Institute for Materials Science in partnership with Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

development that free human researchers from repetitive tasks, through combining automation and artificial intelligence (AI).

In recent years, driven by technological innovations in robotics and artificial intelligence (AI), automated and autonomous experiments have already been implemented in the development of organic and inorganic functional materials that handle liquids [1–3], powders [4–6], thin films [7–9], etc. One exemplar study could be the autonomous experimental system built by Shimizu et al. [8]. They demonstrated autonomous experiments in synthesis and resistance minimization of thin films as an example to illustrate how the speed and the volume of future big-data collection in materials science will be achieved. Another example of the closed-loop autonomous materials exploration could be the work presented by Liang et al. [9], where the experimental determination of phase boundaries in thin films is autonomously interspersed with real-time updating of phase diagram prediction using CALPHAD. Furthermore, orchestration software suites for autonomous experimentation systems and self-driving labs are also being developed to orchestrate experimental actions and data handoff between lab equipment, analysis routines, and experimental planning modules, such as the ARES OS 2.0 [10], ChemOS 2.0 [11] and NIMO [12]. In fact, early autonomous experimentation efforts have focused almost exclusively on functional materials as of 2021, summarized by Stach et al. [13].

Compared with the vast development of automated autonomous materials exploration systems in functional materials, however, systems that handle bulk metallic materials with defined shapes are relatively few. As pointed out by Miracle et al. [14], developing automated autonomous experimental systems for structural materials is identified as an underserved topic due to the additional technical challenges offered by this class of materials. The computational cost is higher, unlike functional materials where the properties of interest can be achieved by DFT calculations; the thermomechanical processing of structural materials and subsequent sample preparation are time-consuming and difficult to automate, and structural characterization and mechanical property evaluation often require different instruments, while specimen-handling procedures are challenging [14]. Existing efforts related to structural metallic materials can be broadly classified into three categories: (i) automated characterization and property evaluation, (ii) automated synthesis of bulk metallic materials, and (iii) data- or AI-driven frameworks for materials design and experimental planning.

First, systems centered on automated characterization and property evaluation of structural materials have recently been developed. Hoefler et al. [15] proposed an automated system for high-throughput process–structure–property dataset generation of structural materials using a temperature gradient heating furnace, automated scanning electron microscopy, and nanoindentation inverse analysis. Although this system enables efficient dataset generation, the applicable processing conditions are constrained by the use of the temperature gradient furnace. Hufnagel et al. [16] developed an automated laboratory integrating a laser microflyer impact system, X-ray diffraction and fluorescence measurements, and nanoindentation for the high-throughput characterization of structural materials for extreme environments. However, the present configuration does not include a process for actively modifying the microstructure, such as heat treatment, and primarily employs thin-foil specimens rather than defined-shape bulk specimens.

Second, other platforms have focused on the automated synthesis of bulk metallic materials. The high-throughput rapid experimental alloy development methodology proposed by Vecchio et al. [17] incorporates automated additive manufacturing synthesis with a revolving powder feeder so that 16 different samples can be produced at once. Further X-ray diffraction, scanning electron microscope (SEM) microstructure analysis and nanoindentation measurement can be performed in an automated, high-throughput fashion. However, sample transfer between the individual instruments is currently performed manually. Another recent example could be the automated synthesis platform for bulk intermetallic compounds based on an arc-melting process developed by Wang et al. [18], where ingots with arbitrary composition can be automatically produced. However, currently, their synthesis still starts from powders or small chunks, and incorporating characterization hardware such as an X-ray diffractometer (XRD) is still underway. These systems provide important capabilities for compositional exploration and alloy synthesis, but they do not automate the complete post-synthesis sequence of processing, microstructural characterization, and mechanical property evaluation for fixed-geometry specimens.

Third, data- and AI-driven frameworks have been developed for autonomous alloy design, experimental planning, data integration, and optimization [19–22]. General orchestration software suites, including ARES OS, ChemOS, and NIMO [10–12], similarly provide functions for coordinating experimental actions and transferring data between instruments, analysis routines, and machine learning-enhanced physics models. These approaches mainly provide the data-management, or orchestration layer and do not, by themselves, establish a complete physical workflow for the automation of experiments.

Taken together, the previous studies have automated important individual stages of structural-materials research. However, to the best of our knowledge, a system that integrates heat treatment,

microstructural characterization, and mechanical property evaluation of defined-shape bulk metallic specimens within a single automated workflow has not yet been reported. Such integration is particularly important for investigating the process–structure–property relationships of commercial structural alloys, for which properties can be extensively varied through processing conditions even without changing the alloy composition. In fact, an almost infinite number of variations could exist when combining the heating temperature, holding time, and heating and cooling rates. Regarding the tempering of martensitic low-alloy steels, the effectiveness of unconventional heat treatment processes such as rapid heating [23, 24] has been proven to improve the mechanical properties (hardness, strength, toughness, etc.) of the alloy. Similarly, the application of multiple steps of tempering should also be promising, but it is seemingly less studied. One report [25] of two-step tempering shows an improvement of the yield-to-tensile-strength ratio. However, the underlying metallurgical mechanisms and optimized conditions for such two-step tempering are still unknown. Considering the potential of microstructure modification and mechanical properties improvement through such unexplored heat treatment, there is a strong need to realize high efficiency and comprehensive exploration across a wide range of parameters.

Accordingly, the aim of the present study is to design and establish an automated experimental system for evaluating the process–structure–property relationships of structural metallic materials. The proposed system integrates an induction heating furnace, an X-ray diffractometer, and an impact testing machine through robotic specimen transfer using a robot arm and unified software control. It thereby enables heat treatment, microstructural characterization, and mechanical property evaluation of defined-shape bulk metallic specimens within a single automated workflow. NIMO is implemented for automated optimization of the furnace PID parameters and provides a basis for future closed-loop autonomous exploration of heat-treatment conditions.

2. System configuration and components

The current system is designed to run fully automatically and is capable of exploring the process–structure–property relationship by performing heat treatments, X-ray diffraction and impact testing on any structural metallic materials with a particular composition, e.g. low-alloy steels. The size of the bulk sample used for this system is fixed at $40 \times 4 \times 1.5\ mm^3$. The schematic diagram of the system configuration is shown in Figure 1. The specimen is heat-treated by an induction heating furnace (Yonekura Manufacturing Automated Autonomous Heating System, custom-made) at a given set of process conditions. The induction furnace is equipped with a vacuum pump, as well as mass flow controllers for introducing an $Ar + 2\%H_2$ mixture gas for the heating atmosphere (to prevent oxidation) and helium gas for gas cooling. It is capable of a heating rate of 500 K/s to 1500 °C and a cooling rate of approximately 200 K/s using helium gas cooling at maximum. Furthermore, all functions of this equipment – including the transport of the sample stage into the chamber, vacuum pumping, gas introduction (including flow rate

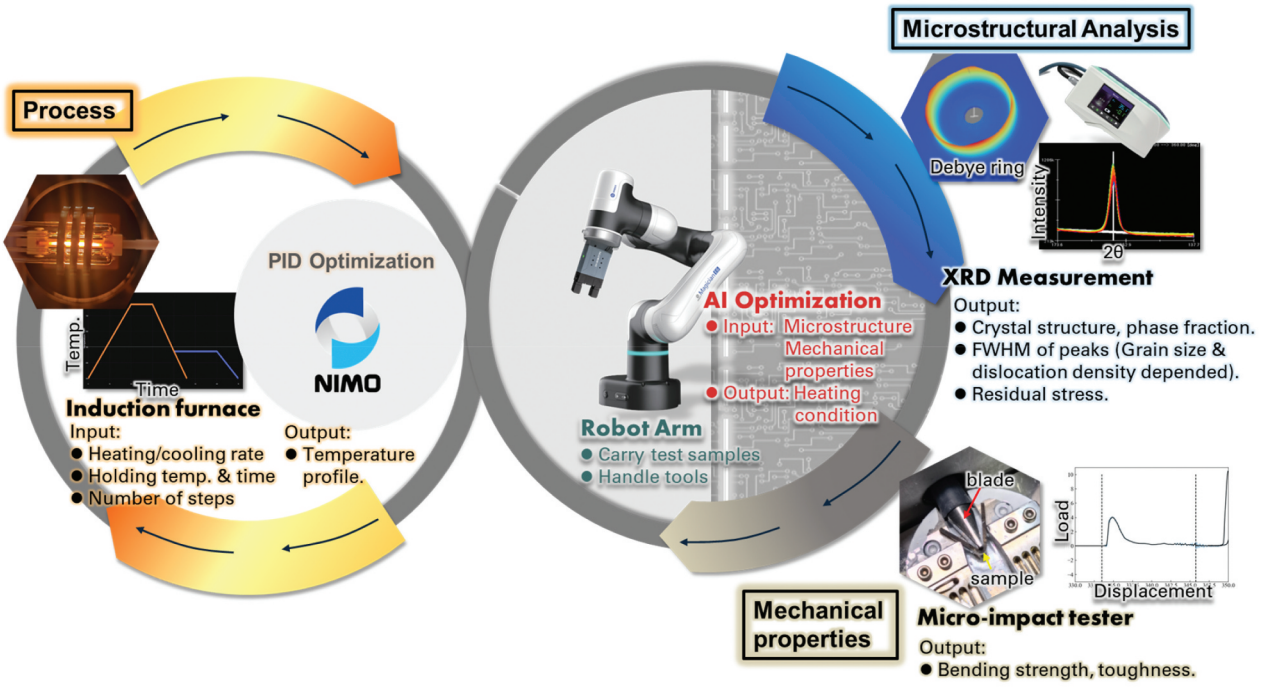


Figure 1. Schematic diagram showing the configuration of the present system.

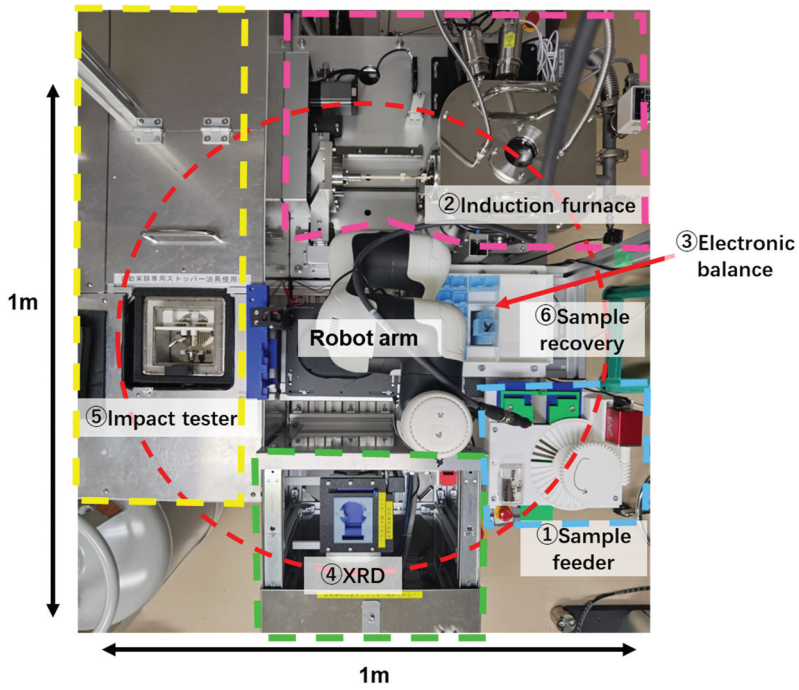


Figure 2. The top view of the entire automated experimental system. The numbers indicate the sequence of the experimental procedure. The red dashed circle indicates the working radius of the robot arm, 450 mm.

control), and the execution of set heating programs – can be automatically controlled using a Python program that sends text-based commands. The temperature of the heated sample is measured and controlled by single-color (150–1000 °C) and two-color (600–1800 °C) infrared pyrometers, each with a measurement spot diameter of approximately 2.5 mm. The homogenous heating zone of the induction coil is approximately 10 mm wide and is sufficient given the X-ray beam diameter (2 mm) and the blade thickness (1 mm at the tip) of the impact testing machine used in this system.

The microstructures of the heat-treated samples are then characterized using a portable X-ray analyser (PULSTEC μ-X360s). The X-ray diffractometer (XRD) is equipped with a 2-dimensional detector and placed in a shielding box with a motorized door. The X-ray for the current diffraction device is produced by a Cr Kα target (Kα complete line (Kα1 + Kα2): 0.2290 nm) with a beam diameter of 2 mm. The X-ray tube is operated at a tube voltage of 30 kV and a tube current of 1.0 mA. It is capable of measuring diffraction peaks of both ferrite (211_{α}, located near $2\theta = 156.4°$) and austenite (220_{γ}, located near $2\theta = 128.8°$) phases for iron-based alloys, thus outputting results on the full width at half maximum (FWHM) and lattice constant of the diffraction peaks, along with phase fractions for the materials. All the diffraction profiles obtained in this study are analyzed using the dedicated software of the analyzer, in which the manufacturer-recommended values are adopted for all the analysis parameters. A one-dimensional diffraction profile is first extracted for each azimuthal angle on the detector and is smoothed by successively applying a median, a Gaussian and a moving-average filter. The background is subsequently evaluated from the whole profile data and subtracted, and the peak position and the peak profile are then determined by fitting a Lorentzian function to the background-subtracted profile, from which the FWHM is obtained. An intensity correction is additionally implemented in the software, and the FWHM contained in the output data is therefore a corrected value; the detailed formulation of this correction is proprietary to the manufacturer and not disclosed. Since an identical instrument, identical measurement conditions and an identical analysis procedure are applied to all the specimens examined in this study, the obtained FWHM values are mutually comparable, and the discussion below is therefore based on their relative changes. Regarding the mechanical properties, a micro-impact tester (TANAKA MIT-D05kJ) is attached to the system, and an impact speed of 1 m/s is adopted. By recording the load vs. displacement curves during impact testing, the bending strength and the impact energies of the materials can be determined as the highest load during deformation and the area under the load–displacement curve, respectively, as explained later. A 6-axis collaborative robot arm (DOBOT Magician E6) is used to carry and place the samples in each part of the system, and its electric gripper allows it to handle different tools for conducting the fully automated experiments.

The top-view photo of the entire automated experimental system is shown in Figure 2. The 6-axis robot arm is placed in the centre of the system. Other

experimental components are placed next to the base of the robot arm. The system is miniaturized within an area of approximately 1 m^2, where the sample stage of the induction heating furnace, the XRD and the testing area of the impact testing machine are all located within the robot arm's reach ($r = 450$ mm). The sequence of the experimental procedure is indicated by the numbers shown in Figure 2. The experiment starts from picking up a new sample from the automated sampling feeder; the heat-treatment is then applied to the sample, and its weight is measured using an electronic balance. The weight of the sample is used to determine whether the collection of the fractured samples has been successful or not after the impact testing. The microstructure of the heat-treated sample is then characterized by XRD. After the X-ray diffraction, the sample is transferred to the impact testing machine for destructive testing of its mechanical properties. The broken samples after impact testing are picked up using an electromagnet gripped by the robot arm and transferred to a separate container placed on the electronic balance. After the weight confirmation, the container is put aside for further processing by human researchers.

Figure 3 shows the screenshots taken from the video recording of the whole experimental process. The full video is shown in Supplementary Movie S1. Corresponding times for the processes represented by the screenshots are also added to Figure 3, where the information is extracted from the experiment logs. An example of the experiment logs for the automated experiment of a single sample is shown in Supplementary Material S2. As shown in Figure 3 and Supplementary Material S2, the approximate duration of each process (excluding the heat treatment time of 3600 s) is summarized as follows: 60 s for sample pick-up from the automated feeder and loading into the furnace, 40 s for sample weighing and transferring to the XRD, 400 s for the X-ray diffraction and analyses, 120 s for the impact testing and analysis, and 180 s for the sample recovery. Therefore, the automated experiments excluding the heat-treatment process can be finished within 800 s.

In the current study, devices and tools are specially designed in order to overcome the difficulties of handling bulk metallic materials during the automated experiments. Some of the devices and tools are shown in Figure 4. Figure 4(a) shows a motorized sample feeding device designed to be built by a 3D printer (Bambu Lab X1C) using polylactic acid filament. It can provide samples to be readily picked up by the robot arm continuously with 40 sample spaces. Driven by a motor-driven gear, each time the holding space aligns with the drop port, a single sample is supplied to the lower conveyor section. This rotating disk-type design allows for sample replenishment by humans without stopping the entire system during automated experiments. For the lower conveyor section, individual samples are transported to the sample pick-up position via the

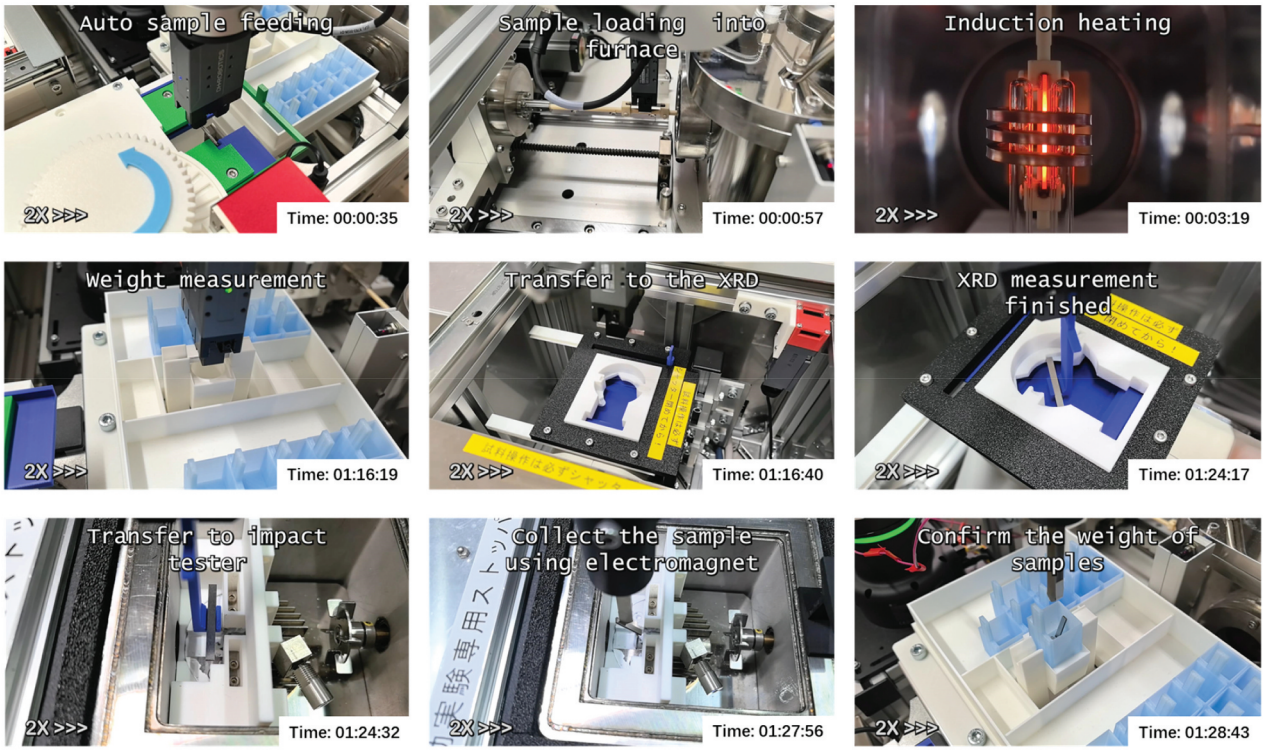


Figure 3. The screenshots taken from the video recording of the entire experimental process. Corresponding time elapsed since the start for each process is also added, with information extracted from the experiment logs. The starting time is 00:00:00, and the finish time is 01:29:38, where an induction heating time of 3600 s is included.

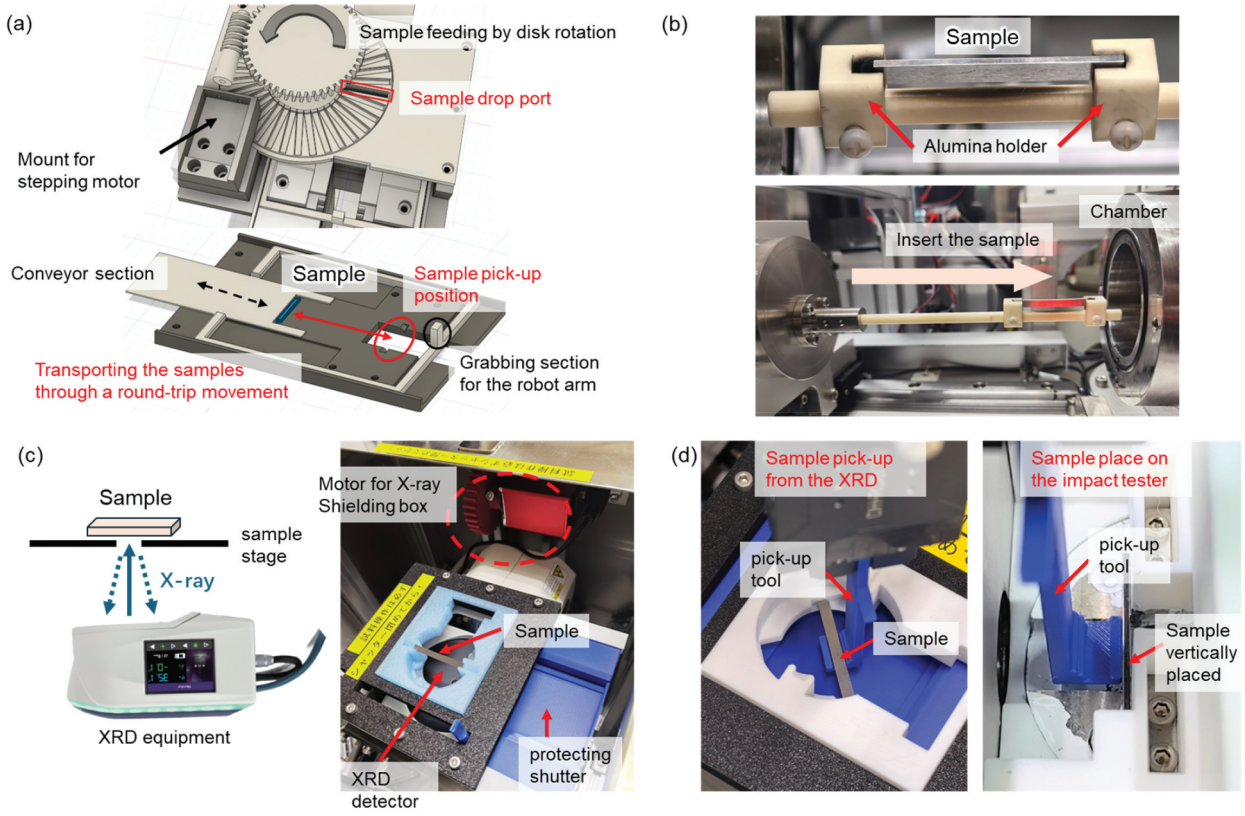


Figure 4. (a) 3D model of the automated sample feeder. (b) Photo of the sample holder used in the induction heating furnace. (c) Schematic diagram and photo of the XRD equipment. (d) Photo of the pick-up tool used for transferring samples between the XRD stage and the impact tester.

lateral reciprocating motion of the white component (powered by the robotic arm), shown in the lower picture of Figure 4(a). Then, they are grabbed by the arm's gripper and moved to the next location. The sample holder for the induction furnace is shown in Figure 4(b). A Y-shaped alumina holder is used to insert the sample into the furnace. The position of the holder is adjusted so that the centre of the sample aligns with the centre of the measurement spot of the single-color infrared pyrometer. For the XRD equipment, the sample is placed from above at the specified camera length, as shown in Figure 4. The 3D-printed sample stage is attached to the XRD, and it features two sample positions: the upper position is designed for residual stress measurement (35° incidence, 44 mm camera length, 211_α peak is measured), and the lower position is designed for retained austenite content measurement (0° incidence, 27 mm camera length, both 211_α and 220_γ peaks are measured). An additional mechanical shutter is designed and positioned between the detector of the XRD and the sample mounting stage to protect the sensor unit from incidental falling objects and collisions with the robot arm. To automate the opening and closing of the shielding box, a stepping motor compatible with Python program control is adopted. For transferring samples between the different camera length positions on the XRD stage and from the XRD to the impact tester, a pick-up tool for the robot arm to handle is designed and shown in Figure 4. The tool is used to place the sample onto the specified position on the impact tester with its side ($40 \times 1.5\ mm^2$) facing upward and its front ($40 \times 4\ mm^2$) facing the blade. The detailed procedure for using such a tool is shown in Supplementary Movie S1.

The temperature control of the induction heating furnace adopted in this system utilizes PID control, and the optimal PID values may vary depending on the target temperature and heating rate. Therefore, incorporating the optimized PID parameters according to the different heat-treatment conditions is necessary to avoid any undesirable overshoot or undershoot. In this study, the NIMS (National Institute for Materials Science) Orchestration System (NIMO, a general-purpose Python library with various combinations of AI modules) [12] is utilized to achieve the automated optimization of PID parameters for the induction heating furnace. A list of PID candidates is prepared in advance, and the NIMO proceeds with the optimization by selecting the next candidate to try from the list. Based on the results of preliminary experiments, the search range for P is set to be 3–15 °C with a step of 1, for I is 0–0.8 s with a step of 0.1, and for D is 0–0.2 s with a step of

0.05. This results in approximately 500 combinations in total. The NIMO sequentially collects measured data, evaluates it based on the deviation from the target temperature profile, and updates the PID parameters in the direction that minimizes the deviation. Assuming the time when the temperature reached a certain value (155 °C in this study) as 0 s and the time when the programmed temperature reached the target temperature as t s, the deviation is evaluated as the integral of the absolute difference between the programmed and actual temperatures over 0 to 2t s using the trapezoidal rule. For the current optimization, PHYSBO in NIMO (an optimization tool for PHYSics based on Bayesian Optimization) [26] is used. An example of such PID optimization for a target temperature of 300 °C and the heating rate of 100 K/s is shown in Figure 5. As shown in Figure 5(a), the deviation generally decreases with increasing trial numbers and significantly low deviations are achieved after 18th trial. Figure 5(b–d) shows the actual heating profiles during the optimization process. For the PID settings prior to optimization (Figure 5(b)), the profile deviated significantly from the target heating profile, and an obvious overshoot is observed. After the search began, the deviation gradually converged toward a smaller value. As shown in Figure 5, by the 18th iteration after optimization began, the actual profile achieved becomes obviously closer to the target ideal profile. Although the optimization was repeated up to 100 iterations, there was virtually no difference between the profile showing the global minimum (25th iteration) and the 18th iteration. Therefore, it is considered that the optimal solution is efficiently reached in approximately 18 iterations, or 1 h. By using this workflow, the optimized PID parameters for arbitrary target temperatures and heating rates can be achieved in the current system.

Figure 6 shows an example of the data achieved by the XRD equipment (0° incidence) and the impact tester adopted in the current system. A quenched and tempered low-alloy steel after mirror polish is used for demonstration. As shown in Figure 6(a), a Debye ring of 211_α diffraction peaks is achieved. The two-dimensional X-ray detector is divided into 500 parts with different azimuthal angles φ on the detector (step width 0.72°), and the change in X-ray intensity is separately recorded. The FWHMs of 211_α peaks for the different azimuthal angles is then analyzed, and the results are shown in Figure 6(b). The values of FWHM for different φ vary due to the heterogeneity of the tempered martensite microstructure. The average FWHM value of the total 500 data points is 5.3 degrees in this example, and the variation among three repeated measurements of the same sample has been confirmed within 0.1 degrees. For the results of impact testing, the load vs. displacement curve is shown in Figure 6. The raw data are smoothed using the Savitzky–Golay method, with a window width of 10 points (0.2 mm) and a polynomial order of 1. A peak of load is observed right after the blade hits the sample at a displacement of 334 mm. The highest load on the deformation peak is then converted to the bending stress using the following equation.

$$\sigma = \frac{3FL}{2bh^2} \tag{1}$$

Where σ is the three-point bending strength, F is the maximum load during testing, L is the span length, b is the width, and h is the thickness of the sample. A bending stress of 2870 MPa was obtained in this measurement. The calculated bending stress in this study is considered positively correlated but not equal to the standardized three-point bending strength testing (e.g. JIS Z 2248 or ISO 7438) due to the following reason: the impact speed of 1 m/s for the

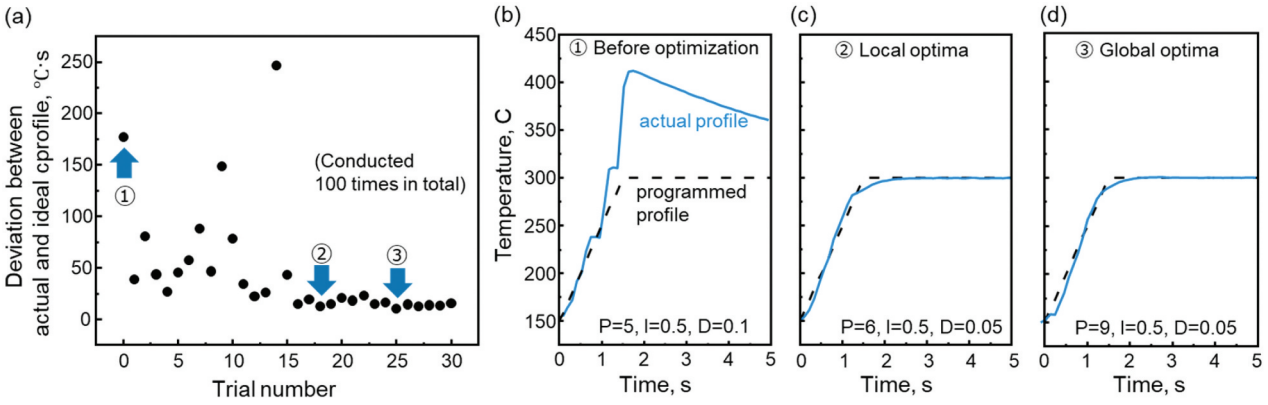


Figure 5. A sample of the autonomous PID parameters optimization by NIMO for the induction heating furnace. (a) Changes in the deviation between the actual and programmed temperature profiles with different trial numbers. Temperature profiles (b) before optimization, (c), (d) achieved for trial numbers 18 and 25, respectively. The target temperature is 300 °C , and the target heating speed is 100 K/s. The overall deviation for each trial shown in (a) is calculated by integrating between the absolute difference between the programmed and actual temperatures using the trapezoidal rule.

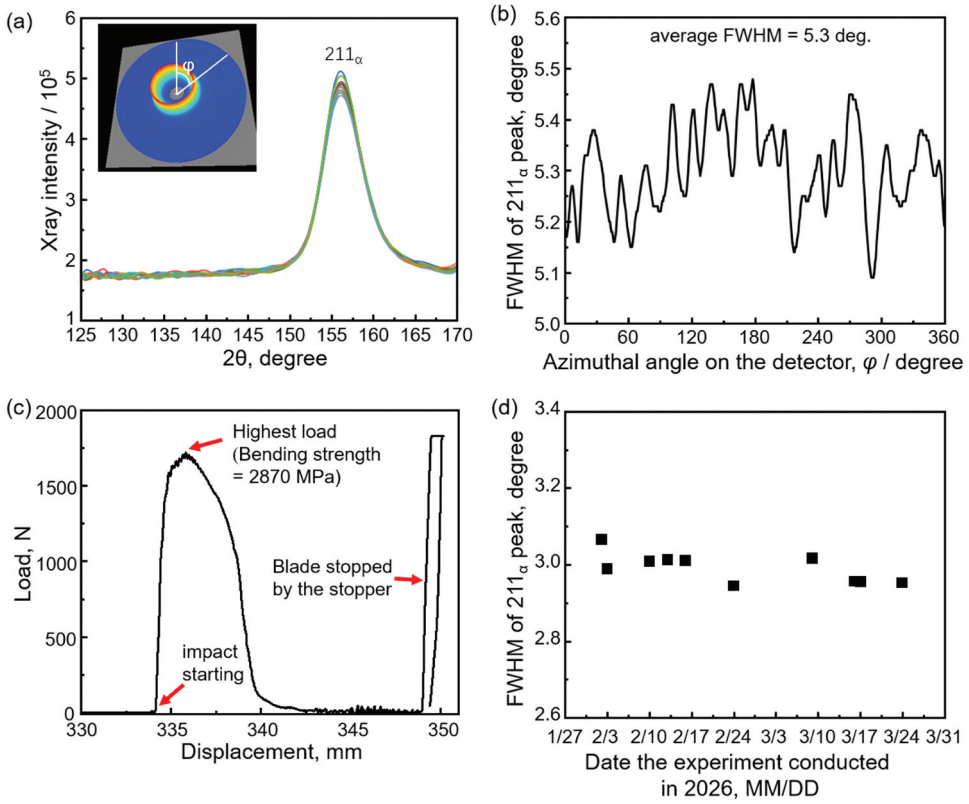


Figure 6. (a) XRD profiles showing 211_α peak for the azimuthal angles of every 30 degrees on the detector; a 3D image of the corresponding Debye ring is also shown in (a) and the azimuthal angle on the detector, φ, is schematically shown. (b) Analyzed FWHM of 211_α peaks for different azimuthal angles on the detector. (c) Load vs. displacement curve for impact testing. The highest load is considered the bending strength and shows a value of 2870 MPa. The tested sample is a JIS SCM435H alloy austenitized at 850 °C for 20 min, water-quenched and then tempered at 300 °C for 30 s. (d) Ring-averaged FWHM of the 211α diffraction peak obtained from 10 repeated automated transport, mounting, and XRD measurement cycles of the same JIS SCM420H reference specimen tempered at 600 °C for 3600 s. The measurements were conducted over a period of approximately 2 months. The size of the error bars is smaller than the size of the data points.

impact tester adopted in the current system is much higher than the standardized tests of 1 mm/s (ISO 7438). Such a high deformation speed was adopted in the current system so that absorption energy can be measured during future impact testing. Nevertheless, the bending stress achieved in the current system is used for the comparisons between the different processing conditions as mentioned later. Furthermore, the absorbed energy during deformation can be calculated by integrating the area enclosed by the load vs. displacement curve. However, the discussion on the absorption energies is not included in this study as currently no notch is added to the samples. A second peak is observed starting from a displacement of 349 mm; this peak corresponds to the stopping of the blade when the impact test is finished.

The accuracy and the reproducibility of the automated XRD measurement were evaluated using a reference specimen of the JIS SCM420H alloy tempered at 600 °C for 3600 s. The specimen was repeatedly picked up, transported, mounted onto the XRD stage and measured by the system, with the identical measurement and analysis conditions described above; the whole sequence was repeated 10 times over a period of approximately 2 months. The ring-averaged FWHM of the 211α peak obtained in each run is plotted in Figure 6(d). The mean value is 2.99 deg with a standard deviation of 0.04 deg, corresponding to a relative standard deviation of 1.3%, and the maximum difference is 0.12 deg. These variations, which include all the uncertainties arising from the repeated transport and mounting operations by the robot arm in addition to those of the X-ray measurement itself, are generally smaller than the change in FWHM induced by the tempering treatments examined in this study as shown later in Figure 7(a). It is therefore concluded that the automated handling sequence does not introduce a significant uncertainty into the measured FWHM, and that the system provides sufficient reproducibility for discussing the effect of the processing conditions. On the other hand, rapid heating and cooling may induce dimensional changes or out-of-plane warping of the strip-shaped specimens, which could affect the effective sample-to-detector distance and diffraction geometry. The current XRD stage with the open-top placement geometry does not actively compensate for substantial local

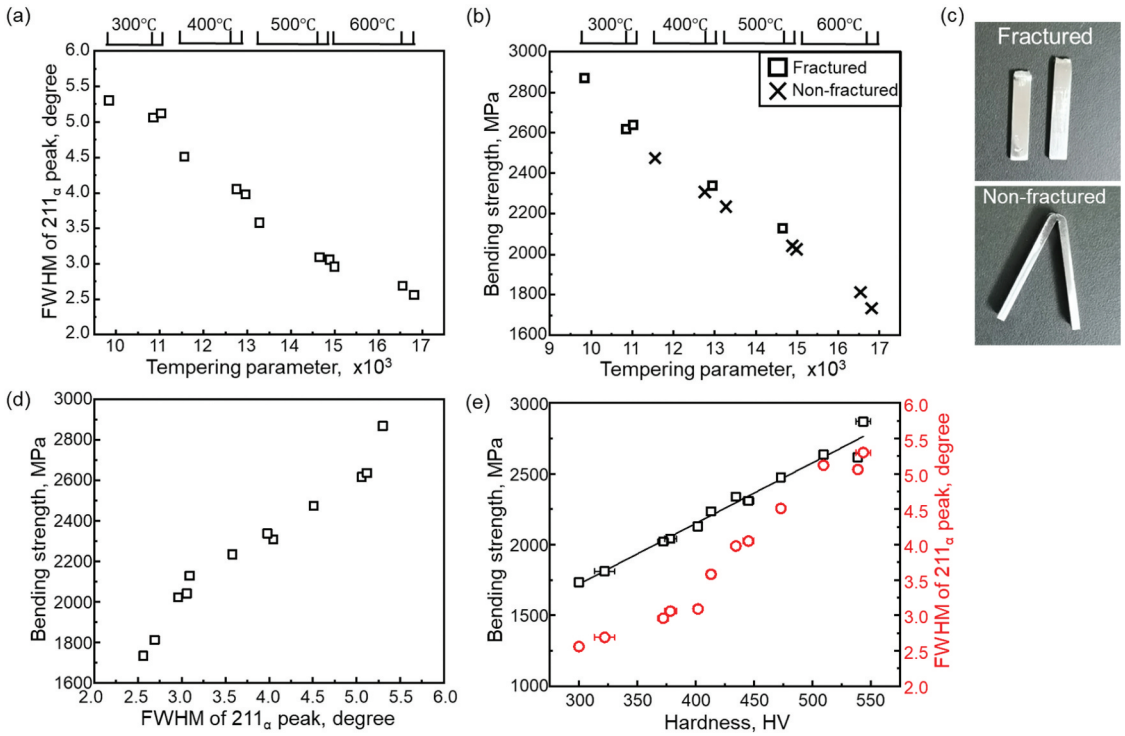


Figure 7. Variations in (a) 211_α peak's FWHM and (b) Bending strength for the one-step tempered SCM435H alloys as a function of tempering parameter. The fracture modes of the one-step samples are indicated in (b) by the □ (fractured) and ✕ (non-fractured) symbols. (c) Photos of the impact-tested samples with different fracture modes. (d) The relationship between the 211_α peak's FWHM and the bending strength for the one-step tempered samples. (e) The variations of the bending strength and the FWHM against the manually measured Vickers hardness of the one-step tempered samples. The black line indicates the linear fitting results of the bending strength over the Vickers hardness, whose R-square value of the fitting result is 0.975.

curvature at the X-ray-irradiated position. For experiments involving higher temperatures or more severe thermal histories, direct specimen-flatness measurements or an additional specimen-height check will be conducted in future work.

Python-based programs are developed to control the robotic arm and integrated devices via TCP/IP connections. The central PC directly controls all devices, except the induction heating furnace and the impact tester, using dedicated Python packages or commands. The induction heating furnace and the impact tester are controlled via text commands sent from the central PC; server programs on their dedicated PCs receive these commands parse them and operate the corresponding software or graphical user interface (GUI) accordingly.

The program is structured into three layers for maintainability and scalability: an orchestration layer, a modular layer and a parameter layer. The orchestration layer defines the sequences of automated experiments, controlling each device through the classes and the parameters in the lower two layers. For PID optimization, the orchestration layer calls NIMO functions to determine the next candidate to evaluate.

The control logic of each device is defined as a dedicated class in the modular layer, separated from the orchestration logic. The parameter layer stores motion orders and kinematic parameters, such as position, speed, and acceleration, for robotic motion. These separations ensure that changes to one device do not affect others and that new devices can be added without modifying existing code. This approach aligns with the SOLID principles [27], a well-established set of guidelines for software design.

Experimental data are uniquely identified by universally unique identifiers (UUIDs) and stored in structured HDF5 [28] files. Metadata, numerical results, logs, and raw data are integrated into a unified format for traceability. This data infrastructure is designed to support future machine learning-based autonomous exploration.

3. Demonstration experiments and results

Demonstration experiments are performed in this study to show the capability and efficiency of the current system. High-throughput automated experiments are performed to investigate the effect of one- and two-step tempering treatments on the mechanical properties of a commercial JIS SCM435H alloy. The precise compositions of the as-received alloy are analyzed using ICP optical emission spectrometry, and the results of alloying elements in mass% are as

follows: 0.37% C, 1.1% Cr, 0.82% Mn, 0.25% Si, 0.16% Mo, 0.027% Al, 0.023% Ni, 0.017% S, and 0.0097% N. The as-received alloys are cut into $80 \times 40 \times 4\ mm^3$ plates and austenitized at 850 °C for 1200 s in a vacuum furnace and quenched in iced water by a human researcher to achieve a full martensitic microstructure. The samples for the automated experiments with a size of $40 \times 4 \times 1.5\ mm^3$ are then cut out by electrical discharge machining. The cut surfaces are mechanically polished using 600 to 1500-grit emery papers. The polished samples are labelled with a laser engraving machine, and their data are managed by their experimental IDs. The tempering conditions for the automated experiments are determined as 300, 400, 500, and 600 °C with a tempering time of 30, 1800 s, and 3600 s for one-step tempering (12 conditions in total), as well as the two-step tempering conditions of 300 °C 30 s + 600 °C 3600 s, 300 °C 3600 s + 600 °C 30 s, 600 °C 30 s + 300 °C 3600 s, and 600 °C 3600 s + 300 °C 30 s. The four two-step tempering conditions were selected as a limited proof-of-capability for automated multi-stage heat treatment rather than as a comprehensive optimization study. They combine two representative tempering temperatures and two holding times in different orders, thereby allowing the influence of the treatment sequence to be examined within a small experimental set. All one-step and two-step samples were processed, characterized, and mechanically tested using the same automated workflow to enable a direct comparison without changing the sample-handling or measurement procedures. Although these four conditions could also be examined by conventional manual experimentation, their automated execution demonstrates a prerequisite for future exploration of substantially larger and more complex multi-step heat-treatment spaces. SEM electron backscatter diffraction (EBSD) characterization of two representative one-step-tempered specimens, 300 °C 3600 s and 600 °C 3600 s, is further manually performed using a field emission SEM (JEOL 7001F) operated at 15 kV with a probe current of 5 nA. The step width of the EBSD was 50 nm, and the results were analyzed using the TSL OIM Analysis 7 software.

In this study, the heat-treatment conditions are categorized based on the tempering parameter (TP) and calculated using the following equations, reported by J.H. Hollomon and L.D. Jaffe [29].

$$TP = T(log_{10}(t) + c) \tag{2}$$

$$c = 21.3 - 5.8 \times mass\%C(for\ t\ in\ h) = 19.2 \rightarrow in\ this\ alloy \tag{3}$$

In Equation 2, *TP* represents the tempering parameter, and *T* represents the absolute temperature. An X-ray exposure time of 15 s is applied for the XRD measurement of the 211_α peak's FWHM. The camera length of the X-ray analyser is calibrated each day so that the measured lattice constant matches the exact value of 0.28664 nm for α phase achieved by a separate XRD analysis, using a standard sample of JIS SCM420H alloy tempered at 600 °C for 3600 s (fully tempered martensitic microstructure). In addition, no retained austenite is detected for all the heat-treated samples in this study. For the verification of the XRD and impact testing results, the microhardness of the one-step heat-treated samples is determined using a Vickers microhardness tester (MATSUZAWA AMT-X7AFS-TU) with a load of 1 kgf.

Figure 7(a,b) shows the measurement results of the 211_α peak's FWHM and bending strength for the one-step tempered SCM435H alloys. The FWHM and the bending strength of the one-step tempered specimens decrease monotonically with increasing TP. The fracture modes of the one-step samples during impact testing are shown in Figure 7(b) and indicated by the □ (fractured) and × (non-fractured) symbols. The samples generally become more ductile with increasing *TP*, and the samples tempered with relatively longer periods do not fracture into two pieces but deform into a V-shape corresponding to the impacting blade. The actual photos of the impact-tested samples with different fracture modes are shown in Figure 7. Figure 7 shows the relationship between the 211_α peak's FWHM and the bending strength. The measured FWHMs show a positive correlation with the bending strength. According to the Williamson–Hall relationship [30], the increase in dislocation density and the decrease in grain size will increase the width of the X-ray diffraction peaks, thus contributing to strength through dislocation strengthening and grain refinement. Since identical measurement and analysis conditions were used for all specimens, the relative changes in FWHM can be compared among the different tempering conditions. Therefore, for the one-step tempered samples in the current study, non-contact, non-destructive strength estimation could be possible with the current system once the relationship between strength and FWHM is established. However, the present measurement of a single diffraction peak does not allow the individual contributions from dislocation structures and microstructural length scales to be quantitatively separated. Finally, Figure 7(e) shows the verification of the accuracy for the bending strength and the FWHM of the 211_α peak. The bending strength measured in the automated experiments correlates linearly with the hardness of the specimens. The black line in Figure 7(e) indicates the linear fitting results of the bending strength over the Vickers hardness, showing a relationship of bending strength = (4.282 ± 0.218)*hardness + (436.11 ± 94.55) MPa, with an R-square value of 0.975. Furthermore, a positive correlation is observed between the

FWHM of the 211_α and the hardness, similar to the bending strength. Therefore, it is clear that the current system enables strength evaluation with a level of accuracy comparable to that achieved by human measurement.

Figure 8 shows the representative SEM-EBSD results of the specimens tempered at 300 °C and 600 °C for 3600 s for the comparison of the tempered martensitic microstructures. According to the EBSD results, the 600 °C 3600 s tempered sample shows a relatively larger average grain size, and a lower value of kernel average misorientation (KAM) value corresponds to a lower dislocation density in the microstructure. Such a difference in the tempered microstructure is directly related to the lower FWHM measured by the XRD. Further characterization using multi-peak X-ray line-profile analysis, EBSD, and TEM will be required to quantitatively determine the respective contributions of dislocation density, crystallographic length scale, recovery, and carbide precipitation, but it is beyond the scope of the current study.

For the automated experimental results of the two-step tempered samples, Figure 9(a) shows the comparison of their bending strength against the corresponding one-step samples. Regarding the TP for a two-step tempering process, it is calculated using the method mentioned in [31], where the heat treatment for

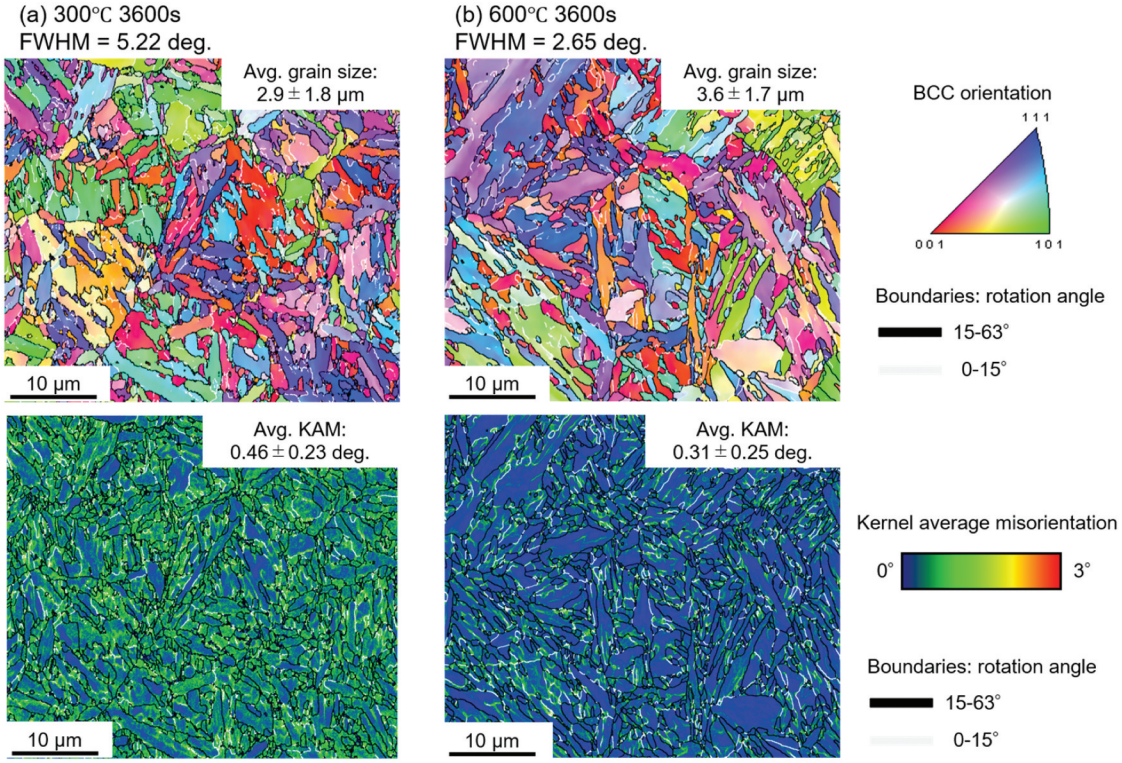


Figure 8. SEM-EBSD results of two representative one-step-tempered specimens, (a) at 300 °C for 3600 s, and (b) at 600 °C for 30 s.

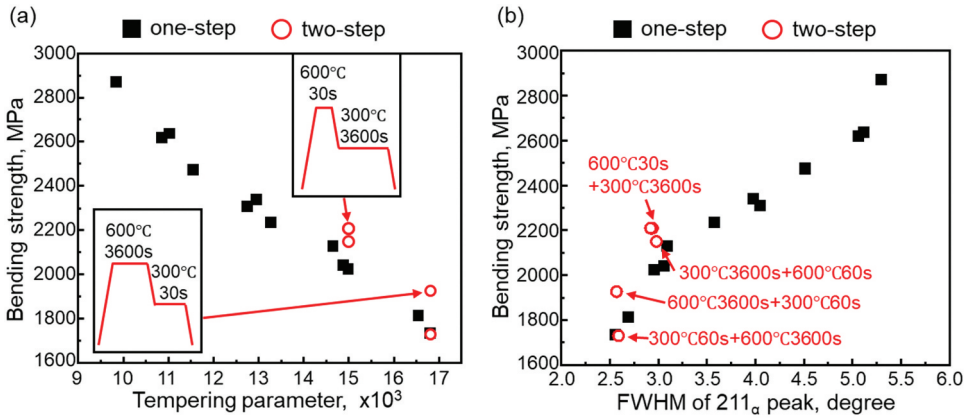


Figure 9. Comparison between one-step and two-step tempered samples: (a) relationship between bending strength and tempering parameter, and (b) relationship between the FWHM of the 211_α peak and the bending strength. The schematic diagrams of the two-step conditions of 600 °C 30 s + 300 °C 3600 s and 600 °C 3600 s + 300 °C 30 s are plotted in (a).

the second step is converted to the effective time at the processing temperature of the first step. For the JIS SCM435 alloy, it is reported that the hardness decreases monotonically with increasing tempering temperature and time for the one-step tempering [23], and no reports of secondary hardening can be found. In the current study, however, compared with the relationship between the bending strength and TP for the one-step samples, the two-step samples of 600 °C 30 s + 300 °C 3600 s, 300 °C 3600 s + 600 °C 30 s and 600 °C 3600 s + 300 °C 30 s show a deviation towards higher strength with an amount of 10–20 MPa, while the condition of 300 °C 30 s + 600 °C 3600 s shows almost no difference after the two-step tempering. The condition of 600 °C 30 s + 300 °C 3600 s shows the highest hardening in comparison with the one-step tempering, although naturally its strength should be lower than the 600 °C 30 s one-step samples due to its longer tempering time. This unusual phenomenon of secondary strengthening induced by two-step tempering has rarely been reported, but revealing its detailed strengthening mechanism is beyond the scope of the current study. The automated experimental results of bending strength and the FWHM of the 211_{α} for the two-step tempered samples are plotted against the one-step tempered samples, as shown in Figure 9(b). Compared with the one-step samples, in which the measured FWHM is positively correlated with the bending strength along a single curve, some of the two-step samples exhibit a clear deviation from the trend and exhibit higher bending strength. The observed excess strengthening of the two-step samples appears undetectable to the current XRD. This indicates that a different balance between dislocation strengthening and grain refinement or additional precipitation hardening may exist. Additionally, the load–displacement curves for the one-step and two-step samples in the current study are shown in Supplementary Material S3. For the two-step samples of 300 °C 3600 s + 600 °C 30 s and 600 °C 30 s + 300 °C 3600 s, the deformation stops earlier during the impact test before the impact blade hits the stopper. This means that the absorption energy during deformation for the two-step samples is significantly improved compared to the one-step samples and is beyond the load ability of the current impact tester. Therefore, the two-step heat treatments performed in this study could have the potential to improve the strength-to-toughness balance of the tempered alloys although the detailed investigation of the underlying mechanism remains a task in future.

The automated experiment system for structural metallic materials developed in the current study can perform one full set of experiments for a single sample within 800 s (excluding the heat treatment). This value should be distinguished from the effective steady-state throughput of the overall workflow, as it is inherently limited by the longer duration of the heat-treatment process. Scheduling optimization and intermediate sample buffering can improve equipment utilization by allowing the heat treatment of one sample to overlap with the characterization, mechanical testing, or recovery of another sample. To support such operation, the software framework is being updated for the parallel processing of multiple samples, and three strategically positioned sample-buffering stages have already been installed in the system. Nevertheless, because the current compact configuration contains only one furnace, one XRD unit, one impact tester, and one robot arm, operational overlap alone cannot eliminate the serial bottlenecks. For the long-holding-time conditions considered in the current study, the achievable reduction in the effective cycle time through such overlap is expected to be only on the order of 10 min per sample. Substantial further increases in throughput would therefore require hardware parallelization, particularly the installation of additional furnaces and, depending on the experimental conditions, additional characterization or testing units. Such expansion would need to be balanced against the compact footprint and cost advantages of the present platform. Future work will therefore combine algorithm-based scheduling with sample buffering to improve resource utilization within the current configuration, while also evaluating hardware parallelization for higher-throughput autonomous exploration.

The current system has the capability to conduct autonomous exploration of the process–structure–property relationship in the near future. The autonomous closed-loop cycle will become possible with the help of NIMO (AI optimization). At a given processing condition, performing PID optimization, and conducting heat treatment, the microstructural information and mechanical properties can be automatically achieved, and the next heating conditions to be explored will be provided by the NIMO based on that information. To further pursue the optimized two-step tempering process, it is possible to perform Bayesian optimization of the two temperatures and two heat-treatment times, targeting an optimized improvement over the mechanical properties relationship obtained for the one-step treatments. Such experiments for the optimization of the two-step tempering process are now in progress, and results will be included in a future publication. Furthermore, since structural information is obtained during the optimization, it is possible to discuss the unique characteristics of two-step heat treatment in terms of the relationship between process conditions and structure, as well as the relationship between structure and properties.

4. Conclusion

In the present study, an automated experimental system for the process–structure–property relationship

evaluation of structural metallic materials is designed and established. The automated experiments are realized through integrating a robot arm with an induction heating furnace, an X-ray diffractometer, and an impact testing machine. Devices and tools for a robot arm to handle, such as an automated sampling feeder, are specially designed in order to overcome the difficulties of handling bulk samples. The characterization process for XRD and mechanical testing can be finished within 800 s for 1 sample. The automated autonomous optimization of PID parameters for the induction heating furnace is realized with the help of NIMO. Demonstration experiments on the tempering of a JIS SCM435H alloy are performed. For the one-step tempered samples with different tempering parameters, the achieved data on 211_{α} peak's FWHM and the bending strength positively correlate with each other, and the measured bending strength is linearly proportional to the hardness measured by a human researcher. The limited explorations of four two-step tempering conditions performed in this study confirmed that the system can execute sequence-dependent multi-stage heat treatments and provide exploratory indications that the multi-step tempering may deviate from the process–structure–property relationships established for one-step tempering. Such demonstration motivates future autonomous exploration of a broader multi-step heat-treatment parameter space. The current system is capable of realizing autonomous exploration in the future by incorporating AI optimization, and the exploration of multi-step tempering of steels will be conducted.

Acknowledgements

The author is indebted to Mr. Fuyuki Sakamoto (Analytical Research Core for Advanced Materials, Institute for Materials Research, Tohoku University) for ICP-AES analysis.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by MEXT Program Data Creation and Utilization Type Material Research and Development Project JPMXP1122684766 and JPMXP1121467561, JSPS KAKENHI Grant-in-Aid for JSPS Fellows [Grant Number 24KJ0367] and JST SPRING [JPMJSP2114], JST FOREST [JPMJFR203W].

ORCID

Yulin Xie http://orcid.org/0000-0002-6555-7171

References

[1] Bédard AC, Adamo A, Aroh KC, et al. Reconfigurable system for automated optimization of diverse chemical reactions. Science. 2018;361(6408):1220–1225. doi: 10.1126/science.aat0650

[2] Grizou J, Points LJ, Sharma A, et al. A curious formulation robot enables the discovery of a novel protocell behavior. Sci Adv. 2020;6(5):eaay4237. doi: 10.1126/sciadv.aay4237

[3] Burger B, Maffettone PM, Gusev VV, et al. A mobile robotic chemist. Nature. 2020;583():237–241. doi: 10.1038/s41586-020-2442-2

[4] Szymanski NJ, Rendy B, Fei Y, et al. An autonomous laboratory for the accelerated synthesis of novel materials. Nature. 2023;624(7990):86–91. doi: 10.1038/s41586-023-06734-w

[5] Chen J, Cross SR, Miara LJ, et al. Navigating phase diagram complexity to guide robotic inorganic materials synthesis. Nat Synth. 2024;3():606–614. doi: 10.1038/s44160-024-00502-y

[6] Ha T, Lee D, Kwon Y, et al. AI-driven robotic chemist for autonomous synthesis of organic molecules. Sci Adv. 2023;9(44):eadj0461. doi: 10.1126/sciadv.adj0461

[7] MacLeod BP, Parlane FG, Morrissey TD, et al. Self-driving laboratory for accelerated discovery of thin-film materials. Sci Adv. 2020;6(20):eaaz8867. doi: 10.1126/sciadv.aaz8867

[8] Shimizu R, Kobayashi S, Watanabe Y, et al. Autonomous materials synthesis by machine learning and robotics. APL Mater. 2020;8(11):111110. doi: 10.1063/5.0020370

[9] Liang H, Wang C, Yu H, et al. Real-time experiment-theory closed-loop interaction for autonomous materials science. Sci Adv. 2025;11(27):eadu7426. doi: 10.1126/sciadv.adu7426

[10] Sloan AWN, Waelder RW, Smith ML, et al. ARES OS 2.0: an orchestration software suite for autonomous experimentation systems and self-driving labs. arXiv preprint. 2026;arXiv:2604.03440. doi: 10.48550/arXiv.2604.03440

[11] Sim M, Vakili MG, Strieth-Kalthoff F, et al. ChemOS 2.0: an orchestration architecture for chemical self-driving laboratories. Matter. 2024;7(9):2959–2977. doi: 10.1016/j.matt.2024.04.022

[12] Tamura R, Tsuda K, Matsuda S. NIMS-OS: an automation software to implement a closed loop between artificial intelligence and robotic experiments in materials science. Sci Technol Adv Mater. 2023;3(1):2232297. doi: 10.1080/27660400.2023

[13] Stach E, DeCost B, Kusne AG, et al. Autonomous experimentation systems for materials development: a community perspective. Matter. 2021;4(9):2702–2726. doi: 10.1016/j.matt.2021.06.036

[14] Miracle DB, Thoma DJ. Autonomous research and development of structural materials – an introduction and vision. Curr Opin Solid State Mater Sci. 2024;33:101188. doi: 10.1016/j.cossms.2024.101188

[15] Hoefler T, Ikeda A, Osada T, et al. Automated system for high-throughput process-structure-property dataset generation of structural materials: a γ/γ′ superalloy case study. Mater Des. 2025;256:114279. doi: 10.1016/j.matdes.2025.114279

[16] Hufnagel TC, Addepalli P, Bhattacharjee A, et al. AIMD-L: an automated laboratory for high-throughput characterization of structural materials

for extreme environments. arXiv preprint. 2026; arXiv:2603.06835. doi: 10.48550/arXiv.2603.06835

[17] Vecchio KS, Dippo OF, Kaufmann KR, et al. High-throughput rapid experimental alloy development (HT-READ. Acta Materialia. 2021;221:117352. doi: 10.1016/j.actamat.2021.117352

[18] Wang WS, Terashima K, Takano Y. Orchestration of heterogeneous experimental machines via ROS2 for automated bulk intermetallic synthesis. Sci Technol Adv Mater: Methods. 2026;6(1):2611510. doi: 10.1080/27660400.2025.2611510

[19] Yang P, Zhao C, Tang B, et al. Autonomous multi-objective alloy design through simulation-guided optimization. arXiv Prepr. 2025;arXiv:2507.16005. doi: 10.48550/arXiv.2507.16005

[20] Ha M, Le D, Nguyen V, et al. Beyond interpolation: integration of data and AI-extracted knowledge for high-entropy alloy discovery. Digit Discov. 2026;5 (3):1079–1097. doi: 10.1039/D5DD00400D

[21] Shi T, Li Y, Wang Z, et al. Knowledge-driven autonomous materials research via collaborative multi-agent and robotic system. Matter. 2026;9 (2):102577. doi: 10.1016/j.matt.2025.102577

[22] Mulukutla M, Person AN, Voigt S, et al. Illustrating an effective workflow for accelerated materials discovery. Integr Mater Manuf Innov. 2024;13 ():453–473. doi: 10.1007/s40192-024-00357-3

[23] Furuhara T, Kobayashi K, Maki T. Control of cementite precipitation in lath martensite by rapid heating and tempering. ISIJ Int. 2004;44(11):1937–1944. doi: 10.2355/isijinternational.44.1937

[24] Judge VK, Speer JG, Clarke KD, et al. Rapid thermal processing to enhance steel toughness. Sci Rep. 2018;8 ():445. doi: 10.1038/s41598-017-18917-3

[25] Yoshioka Y, Itou H, Tanaka Y, et al. The influence of chemical composition and temper heat-treatment on strength and impact toughness in Ni-Cr-mo-V steels. Tetsu-To-Hagane. 2003;89(6):705–710. doi: 10.2355/tetsutohagane1955.89.6_705

[26] Motoyama Y, Tamura R, Yoshimi K, et al. Bayesian optimization package: PHYSBO. Comput Phys Commun. 2022;278:108405. doi: 10.1016/j.cpc.2022.108405

[27] Martin RC. Agile software development: principles, patterns, and practices. Upper Saddle River NJ: Prentice Hall; 2002.

[28] The HDF Group. Hierarchical data format, version 5 (HDF5). 1997–2025. Available from: https://www.hdfgroup.org/HDF5/

[29] Hollomon JH, Jaffe LD. Time-temperature relations in tempering steel. Trans AIME. 1945;162:223–249.

[30] Williamson GK, Hall WH. X-ray line broadening from filed aluminium and wolfram. Acta Metallurgica. 1953;1():22–31. doi: 10.1016/0001-6160(53)90006-6

[31] Inoue T. A New tempering parameter and its application to the integration of tempering effect of continuous heat cycle. Tetsu-To-Hagane. 1980;66(10):1532–154. doi: 10.2355/tetsutohagane1955.66.10_1532