Image-Based Machine Learning Using Inkjet-Printed Chemicals: Mixing Ratio Prediction and Metal Ion Detection

Taichi Sano,[a] Yuki Terauchi,[a] Yuki Ide, *[,b] Ichigaku Takigawa, *[,b,c,d] Tsuyoshi Minami,*[,e] and Yasuhide Inokuma*[,a,b]

[a] Division of Applied Chemistry, Faculty of Engineering, Hokkaido University, Kita 13, Nishi 8, Kita-ku, Sapporo, Hokkaido 060-8628, Japan.

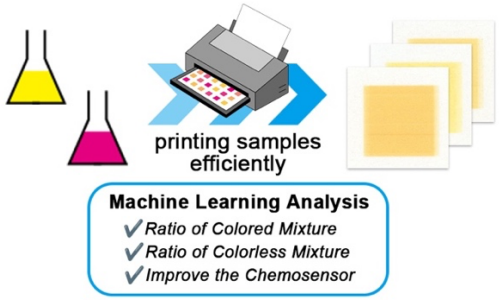
[b] Institute for Chemical Reaction Design and Discovery (WPI-ICReDD), Hokkaido University, Kita 21, Nishi 10, Kita-ku, Sapporo, Hokkaido 001-0021, Japan.

[c] Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-8561, Japan.

[d] RIKEN Center for Advanced Intelligence Project, 1-4-1 Nihonbashi, Chuo-ku, Tokyo 103-0027, Japan.

[e] Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan.

Supporting Information Placeholder



ABSTRACT: Inkjet printing of π-conjugated organic compounds enabled rapid, low-cost generation of training images for image-based machine learning (ML) prediction of mixing ratios. ML models with mean absolute errors below 4% were achieved within hours even for dyes with subtle color differences. Changing the printing surface from filter paper to polypropylene film extended the method to colorless compounds, including isomeric and macrocyclic systems. This approach also enabled spatial mapping of sub-microgram levels of Zn^{2+} ions using a weakly responsive colorimetric sensor, without the need for a spectrometer. This work demonstrates a simple, versatile strategy for integrating π-conjugated materials with ML in colorimetric sensing and mixture analysis.

Machine learning (ML) has become an indispensable tool in chemical science, contributing to data analysis,[1] property prediction,[2] and automation of synthesis.[3] While various ML architectures are available, model accuracy largely depends on the quality and quantity of training data. In chemistry, such data are typically derived from experiments or theoretical calculations. When large datasets are needed for high-performance models, ML predictions may offer limited practical benefit. Although high-throughput equipment can generate extensive datasets,[4] their cost remains a barrier. Thus, developing cost-effective methods to acquire sufficient training data is essential for broader ML adoption in data-driven chemistry.

We recently developed an image-based ML system capable of predicting the mixing ratios of various compounds, including organic/inorganic, crystal polymorphs, and enantiomeric mixtures, with reasonable mean absolute errors (MAEs) of ~5%.[5] This system has also been applied to monitoring solid-state reactions by predicting yields and product compositions.[6] While this non-destructive, equipment-free method is promising, the training data acquisition remains labor-intensive. In the prototype, model construction required ~300 images, each generated from 200 mg of sample, and a new model was needed for each compound or reaction. The most time-consuming step was the manual preparation of mixtures at different ratios.

Inkjet printers have been widely used in chemical research to deposit small amounts of chemicals at predefined locations, particularly for organic electroluminescent devices,[7] microsensor arrays,[8] and microfluidic systems.[9] Our group has applied this technology to create paper-based sensor arrays for the high-throughput analysis of pH, metal ions, and small molecules.[10] A key advantage of inkjet printing is its ability to rapidly and reproducibly generate multi-component mixtures at varying ratios using minimal sample quantities. Based on this, we addressed the bottoleneck of training data acquisition in image-based ML for mixing ratio prediction. Here, we report an efficient protocol to generate up to 260 images on an A4-sized sheet, each containing 6.0~100 μg of sample, using a commercial inkjet printer. A pilot study with π-conjugated chromophores, pyrocatechol violet (**1**) and alizarin red S (**2**), printed on filter paper achieved high prediction accuracy (MAE = 3.7%), comparable to

manually prepared datasets. Using polypropylene sheets enabled accurate ML prediction for colorless compounds. The method was also extended to dyes with similar colors, *cis*/*trans* isomers, and macrocyclic systems. Furthermore, combining ML with a π-conjugated metal ion sensor allowed mapping of sub-microgram Zn^{2+} on paper, without a spectrometer.

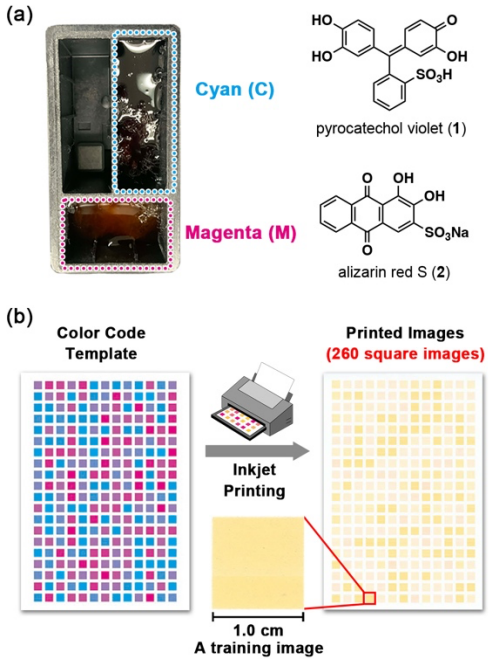


Figure 1. (a) Loading of chromophores **1** and **2** into the cyan and magenta ink slots. (b) Schematic representation of training data generation using a color code template and inkjet printing.

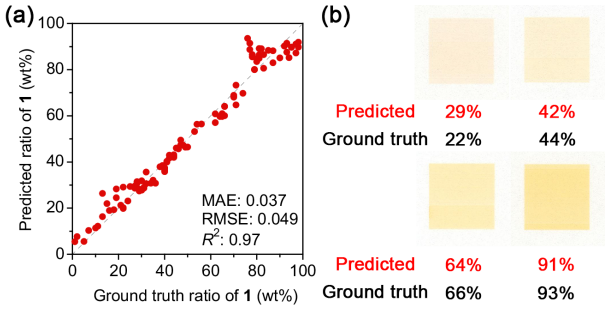


Figure 2. (a) Observation-prediction plot for the weight fraction of chromophore **1** in mixtures with **2**. (b) Representative images with predicted and ground truth values of ratio of **1**, where ground truth was defined by the color template.

An inkjet printing protocol was established to efficiently generate mixtures of two chromophores, **1** and **2**, at varied ratios. Each was dissolved in water (4.00 g/L) and loaded into the cyan (C) and magenta (M) ink slots (Figure 1a) of a commercial printer (Canon iP2700[11]; resolution: 4800 × 1200 dpi). Mixing ratios were controlled via the CMYK color model using a Python-generated color code template (see Supporting Information (SI), Figure 1b, S1, S2). The template consisted of 260 squares (1 × 1 cm) arranged in a 13 × 20 grid on an A4-sized sheet at 600 ppi resolution (Figure 1). The color of each square area was defined by setting the C and M values to x and y, respectively ($0 \le x, y \le 100\%$), representing the mixing weight ratio of **1** to **2** as x/y. When $x + y = 100\%$, the total sample amount in each area was approximately 6.4 μg/cm^2.

Chromophores **1** and **2** were inkjet-printed onto A4-sized filter paper (Whatman Grade 1) with mixing ratios satisfying $x + y = 100\%$. The printed sheet was scanned at 600 dpi using an EPSON GT-X830 scanner, and digital images were saved as TIFF files. Using a Python script (see SI), each of the 260 square was cropped into an individual image named *cropped_image_z.tiff*, where z donates the row-major index. Of these, 100 images were used to train the convolutional neural network (CNN) model for softmax regression, referred to as Model 1 (details are provided in SI), which is a minor modification of our previously reported model.[5] Prediction accuracy was evaluated using another 100 images, and the results are shown in the observation–prediction plot (Figure 2). The MAE, root mean squared error (RMSE), and coefficient of determination (R^2) for predicting the weight ratio of **1** were 3.7%, 4.9%, and 0.97, respectively, comparable to models trained with manually collected data. Importantly, the inkjet protocol was markedly more time- and cost-efficient. While the prototype required several days to collect ~300 training images by weighing each sample on a 200 mg scale, the inkjet method generated up to 260 images per sheet within a few hours.

The versatility of the trained ML model was further evaluated using alternative data preparation methods. Although training images were generated by overprinting **1** and **2** on filter paper, reasonable accuracy was achieved even with pre-mixed solutions (MAE: 7.7%, RMSE: 9.8%, and R^2: 0.85; Figure S8). The model also maintained comparable performance across a sample loading range of 3–7 μg/cm^2, despite being trained at 6.4 μg/cm^2 (Table S1).

While training data for mixing ratio prediction of chromophores were successfully generated by inkjet printing on filter paper (Figure 4a,b), the same approach proved ineffective for colorless compounds. When salicylic acid (**3**) and acetylsalicylic acid (**4**) were inkjet-printed at 10 μg/cm^2, the resulting ML model (trained with 100 images) yielded poor accuracy (MAE = 24%, RMSE = 28%) on 100 test samples, indicating a marked drop in performance. The images of **3** and **4** appeared virtually indistinguishable to the naked eye. To overcome this, we optimized the printing surface for colorless compounds using mixtures of **3** and **4**.

Table 1. Prediction accuracy of ML models for the mixing ratio of **3** in mixtures with **4**, using various printing substrates.

Printing surfaces	MAE	RMSE	R^2
White filter paper (Whatman's Grade 1)	0.24	0.28	–
Black copy paper (thickness: 0.09 mm)	0.25	0.29	–
Black copy paper (thickness: 0.18 mm)	0.24	0.29	–
Polypropylene sheet (thickness: 0.25 mm)	0.085	0.10	0.86
Polypropylene sheet (thickness: 0.30 mm)	0.082	0.10	0.87

When black copy paper was used for inkjet printing, solutions of **3** and **4** penetrated the substrate, producing negligible visible changes, similar to filter paper. ML-based mixing ratio prediction with these images failed. Although a previous report[12]

highlighted the effect of paper thickness on colorimetric analysis, increasing the thickness from 0.09 to 0.18 mm had no effect on prediction accuracy (Table 1). To suppress penetration, mixtures of **3** and **4** were printed on polypropylene sheet (0.25 mm), where colorless solids remained on the surface. Scanning against a black background, an ML model trained with 100 images predicted the ratios with MAE = 8.5%, RMSE = 10%, and R^2 = 0.86. Further increasing the sheet thickness to 0.30 mm showed no improvement, prompting sample amount optimization.

The prediction accuracy of ML models was evaluated based on total sample amounts per square (10–100 μg/cm^2), adjusted by overprint cycles (1, 3, 5, 7, and 10). Each model was trained with 100 images printed on 0.30 mm-thick polypropylene sheets, corresponding to total amounts of 10, 30, 50, 70, and 100 μg/cm^2 for compounds **3** and **4**. MAE and RMSE improved up to 5 cycles, but slightly worsened with further overprinting (Figure 3a). The best accuracy was achieved with 50 μg/cm^2 (MAE = 4.5%, RMSE = 5.7%, and R^2 = 0.96). Although the model tended to underestimate **3** at >90 wt%, it performed well in the 0–70 wt%.

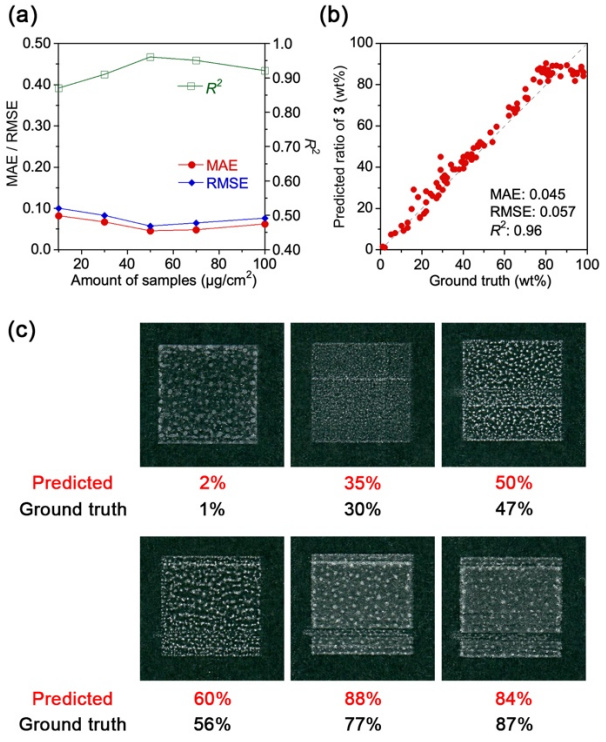


Figure 3. (a) Prediction accuracy for the mixing ratio of salicylic acid **3** in mixtures with **4** as a function of printing cycles. (b) Representative images with predicted and ground truth values.

Based on the pilot studies, we further evaluated the ML system using various compound mixtures (Figure 4). Methyl orange (**5**), which appears similar in color to **1** when inkjet-printed on filter paper, was more difficult to distinguish than mixtures of **1** and **2**. The ML model predicted mixing ratios of **1** and **5** with MAE = 3.9%, RMSE = 5.1%, and R^2 = 0.97 (Figure 4a). 2,2'-Dihydroxyazobenzene (**6**), an azo dye, is known to degrade into colorless *o*-aminophenol (**7**).[13] This dye–degradation product pair also yielded accurate prediction via inkjet-printing (MAE = 4.7%, RMSE = 6.2%, and R^2 = 0.95), suggesting potential utility for monitoring azo dye degradation. Fumaric acid

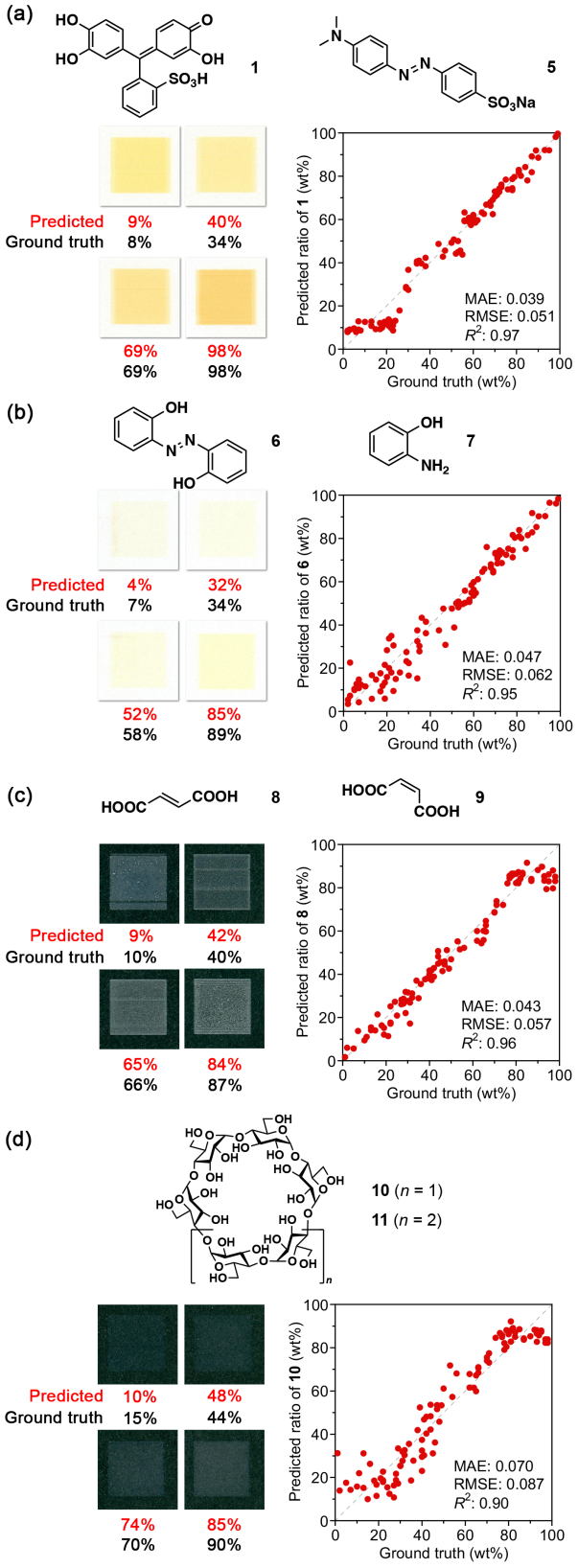


Figure 4. ML Prediction results for component ratio analysis of (a) **1** from mixtures of **1** and **5**, (b) **6** from mixtures of **6** and **7**, and (c) **8** from **8** and **9**, (d) **10** from **10** and **11**.

(**8**) and maleic acid (**9**) are *trans*/*cis* isomers that interconvert via photo-/thermal isomerization.[14] Due to their colorless nature, ML-based mixing ratio prediction was conducted using samples printed on polypropylene. An ML model trained with 100 images (7 overprints, ~70 µg each) achieved MAE = 4.3%, RMSE = 5.7%, and R^2 = 0.96 on 100 test images.

Well-known supramolecular hosts, α- and γ-cyclodextrins (**10** and **11**; containing 6 and 8 glucose units, respectively), showed minimal visual difference when inkjet-printed. An ML model trained under standard conditions (polypropylene, 1 printing cycle, ~10 µg/image) predicted their mixing ratios with MAE = 7.0%, RMSE = 8.7%, and R^2 = 0.90. Prediction errors increased above 80% composition of either component, while higher accuracy was observed in the 30–70% range, likely due to crystallization of cyclodextrins at higher purities altering the appearance of printed solids. When *o*-, *m*-, and *p*-aminophenols (**7**, **12**, and **13**) were printed using the cyan, magenta, and yellow ink slots, respectively, mixing ratio prediction in a three-component system was achieved. An ML model trained with 100 images showed limited accuracy (MAE = 13%), but expanding the training set to 400 images improved the prediction for **7** (MAE = 7.4%, RMSE = 9.8%, and R^2 = 0.92). Predictions for **12** and **13** showed similar accuracy (Figure S23). This result underscores the inkjet protocol's advantage in enabling additional training through rapid, low-cost generation of mixture images.

The image-based ML system expanded the applicability of a metal ion sensor with limited colorimetric response. Chromophore **14** binds Zn^{2+} in solution, causing visible absorption changes.[15] The detection limit by UV-vis spectroscopy in CH_2Cl_2/methanol (v/v 1:1) was 4.6 µM (Figure S7). However, detecting sub-microgram levels of Zn^{2+} with **14** printed on filter paper was difficult due to minimal visible color change. To construct an ML model, methanol/water (v/v 3:1) solutions of **14** (2.0 mM) and $Zn(OAc)_2$•$2H_2O$ (4.0 mM) were loaded into the cyan and magenta ink slots. A total of 100 training images were printed with Zn^{2+} amounts from 0 to 6.2 µg per image (0–2.00 equivalents relative to **14**). These 100 images were used to train the CNN model for positive value regression, referred to as Model 2 (details provided in SI) from our previously reported model.[6]

An observation–prediction plot with 100 test images gave a prediction accuracy of MAE = 0.78 µg/cm^2, RMSE = 1.0 µg/cm^2, and R^2 = 0.64. Since increasing the training set showed minimal improvement, we next examined printing resolution. Lower the resolution enhanced accuracy, with 150 ppi yielding MAE = 0.52 µg/cm^2, RMSE = 0.75 µg/cm^2, and R^2 = 0.84 (Table S10). Previous report noted that bothexcessively high and low resolutions cause inhomogeneous printing.[16] In our case, performance improved down to 150 ppi, but declined at 72 ppi, likely due to suboptimal droplet spacing and mixing. Using 200 training images at 150 ppi, we achieved MAE = 0.39 µg/cm^2, RMSE = 0.55 µg/cm^2, R^2 = 0.90, and a Zn^{2+} detection limit of 0.96 µg/cm^2.[17]

Given the sufficient sensitivity of inkjet-printed **14** to Zn^{2+} ions at the microgram scale, we applied the ML system to visualize Zn^{2+}-distribution on paper. A sensor sheet with 260 squares (5.2 µg/cm^2 of **14**) was printed on A4-sized filter paper. Zn^{2+} ions (600 µg total, 6.0 µg/cm^2 average) was sprayed as a methanol solution of $Zn(OAc)_2 \cdot 2H_2O$ using a mister. After drying, no visible difference was observed between sprayed and non-sprayed areas, but ML-based mapping clearly revealed the Zn^{2+}-distribution, matching the sprayed pattern (Figure 5). Notably, this approach expand the detection range of **14** without requiring an expensive spectrometer.

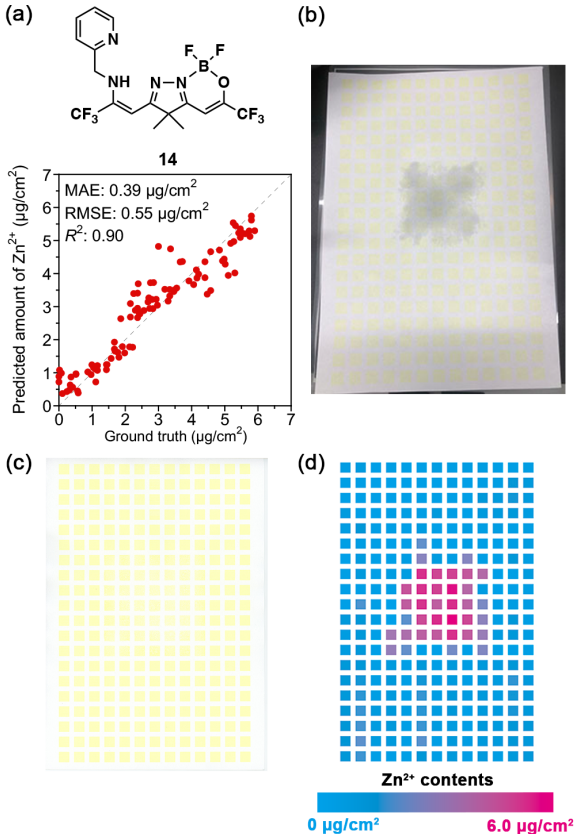


Figure 5. (a) ML-predicted Zn^{2+} content using inkjet-printed sensor **14**. (b) Sensor sheet immediately after Zn^{2+} solution spray. (c) Scanned image after drying. (d) Zn^{2+} distribution map generated by ML.

The inkjet printing protocol enabled cost- and time-efficient generation of training images for image-based ML prediction of mixing ratios. Various π-conjugated chromophores, including dyes with similar colors, were successfully used, achieving MAEs below 8%. By changing the printing surface, the method was extended to colorless compounds. The developed ML models show potential for monitoring *cis*/*trans* isomerization, quantifying macrocyclic ring size selectivity, and analyzing multicomponent mixture. Zn^{2+} ion detection using inkjet-printed sensor **14** demonstrated that ML can extend the utility of sensors with limited sensitivity or subtle colorimetric responses, without requiring expensive spectrometers. Given the wide variety of chromophores with intrinsic properties such as stimuli-responsiveness, guest inclusion, and light emission, this protocol offers a promising bridge between traditional organic chemistry and machine learning. Then, the ML system with inkjet-printed compounds could be applied as a quick and cost-effective analytical tool for reaction yield estimation, trace metal ion detection, and monitoring of chromophore degradation.

ASSOCIATED CONTENT

Data Availability Statement

The data underlying this study are available in the published article, in its Supporting Information , and openly available in figshare at https://doi.org/10.6084/m9.figshare.29194859.v3.

・General information, details of ML, compound images for ML (PDF)
・Model 1, Model 2, utils, and mapping codes (ZIP)

Supporting Information Statement

The Supporting Information is available free of charge on the ACS Publications website. General information, details of ML, compound images for ML (PDF).

・Model 1, Model 2, utils, and mapping codes (ZIP)

AUTHOR INFORMATION

Corresponding Author

* Yuki Ide – Email: ide.yuki@icredd.hokudai.ac.jp
* Ichigaku Takigawa – Email: takigawa@icredd.hokudai.ac.jp
* Tsuyoshi Minami – Email: tminami@g.ecc.u-tokyo.ac.jp
* Yasuhide Inokuma – Email: inokuma@eng.hokudai.ac.jp

Author Contributions

Y. Inokuma designed and supervised the project. T. Minami supervised the use of inkjet printing for chemical compounds. I. Takigawa developed the ML systems. T. Sano, Y. Terauchi, and Y. Ide performed image collection and construction of the datasets. All of the authors discussed the results and contributed to writing the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

This work was partly supported by a JSPS Grant-in-Aid for Challenging Research (Exploratory) (No. JP24K2178704), a JST FOREST Program (No. JPMJFR211H) and the Asahi Glass Foundation, of which Y. Inokuma is the principal investigator. T. Minami gratefully acknowledges financial support from the Japan Society for the Promotion of Science (JSPS) KAKENHI (Grant Nos. JP23H03864 and JP24K01315) and JST CREST (Grant No. JPMJCR2011). This work was partly supported by JSPS KAKENHI (Grant Nos. JP20H05657, JP25H01114, and JP25K03174) to I. Takigawa. The Institute for Chemical Reaction Design and Discovery (ICReDD) was established by World Premier International Research Initiative (WPI), MEXT, Japan.

REFERENCES

(1) Brereton, R. G.; Lloyd, G. R. Support Vector Machines for classification and regression. *Analyst* **2010**, *135*, 230–267.

(2) Muratov, E. N.; Bajorath, J.; Sheridan, R. P.; Tetko, I. V.; Filimonov, D.; Poroikov, V.; Oprea, T. I.; Baskin, I. I.; Varnek, A.; Roitberg, A.; Isayev, O.; Curtalolo, S.; Fourches, D.; Cohen, Y.; Aspuru-Guzik, A.; Winkler, D. A.; Agrafiotis, D.; Cherkasov, A.; Tropsha, A. QSAR without borders. *Chem. Soc. Rev.* **2020**, *49*, 3525–3564.

(3) Coley, C. W.; Eyke, N. S.; Jensen, K. F. Autonomous Discovery in the Chemical Sciences Part I: Progress. *Angew. Chem. Int. Ed.* **2020**, *59*, 22858–22893.

(4) Burger, B.; Maffettone, P. M.; Gusev, V. V.; Aitchison, C. M.; Bai, Y.; Wang, X.; Li, X.; Alston, B. M.; Li, B.; Clowes, R.; Rankin, N.; Harris, B.; Sprick, R. S.; Cooper, A. I. A mobile robotic chemist. *Science* **2020**, *583*, 237–241.

(5) Ide, Y.; Shirakura, H.; Sano, T.; Murugavel, M.; Inaba, Y.; Hu, S.; Takigawa, I.; Inokuma, Y. Machine Learning-Based Analysis of Molar and Enantiomeric Ratios and Reaction Yields Using Images of Solid Mixtures. *Ind. Eng. Chem. Res.* **2023**, *62*, 13790–13798.

(6) Sano, T.; Ide, Y.; Tsumori, T.; Ubukata, H.; Takigawa, I.; Kageyama, H.; Inokuma, Y. Hydride Content Control of Perovskite Oxyhydride $BaTiO_{3-x}H_x$ Supported by Image-Based Machine Learning. *ACS Appl. Eng. Mat.* **2024**, *2*, 2391–2396.

(7) Sirringhaus, H.; Kawase, T.; Friend, R. H.; Shimoda, T.; Inbasekaran, M.; Wu, W.; Woo, W. P. High-Resolution Inkjet Printing of All-Polymer Transistor Circuits. *Science* **2000**, *290*, 2123–2126.

(8) Soga, T.; Jimbo, Y.; Suzuki, K.; Citterio, D. Inkjet-Printed Paper-Based Colorimetric Sensor Array for the Discrimination of Volatile Primary Amines. *Anal. Chem.* **2013**, *85*, 8973–8978.

(9) Martinez, A. W.; Phillips, S. T.; Butte, M. J.; Whitesides, G. M. Patterned Paper as a Platform for Inexpensive, Low-Volume, Portable Bioassays. *Angew. Chem. Int. Ed*. **2007**, 46, 1318–1320.

(10) Mohan, B.; Sasaki, Y.; Minami, T. Paper-based optical sensor arrays for simultaneous detection of multi-targets in aqueous media: A review. *Anal. Chim. Acta.* **2024**, *1313*, 342741.

(11) The same series of inkjet printer is available outside of Japan. For example: Canon PIXMA iP2700 for the Europe. https://www.canon-europe.com/support/consumer/products/printers/pixma/ip-series/pixma-ip2700.html (accessed May 30, 2025).

(12) Pratiwi, R.; Septyani, R. N.; Febriany, R.; Saputri, F. A.; Nuwarda, R. F. Design and Optimization of Colorimetric Paper-Based Analytical Device for Rapid Detection of Allopurinol in Herbal Medicine. *Int. J. Anal. Chem.* **2019**, 4682839.

(13) Alonso, F.; Radivoy, G.; Yus, M. Reduction of Hydrazines, Azo and Azoxy Compounds, and Amine *N*-Oxides with the $NiCl_2$•$2H_2O$–Li–DTBB (cat.) Combination. *Tetrahedron* **2000**, *56*, 8673–8678.

(14) Terryl, E. M.; Eichelberger, L. Catalytic Transmutation of Maleic Acid into Fumaric Acid. *J. Am. Chem. Soc.* **1925**, *47*, 1402–1412.

(15) Shirakura, H.; Manabe, Y.; Kasai, C.; Inaba, Y.; Tsurui, M.; Kitagawa, Y.; Hasegawa, Y.; Yoneda, T.; Ide, Y.; Inokuma, Y. Isopyrazole-Masked Tetraketone: Tautomerism and Functionalization for Fluorescent Metal Ligands. *Eur. J. Org. Chem.* **2021**, *2021*, 4345–4349.

(16) Nasstorom, H.; Shargaieva, O.; Becker, P.; Mathies, F.; Zizak, I.; Schroder, V. R. F.; List-Kratochvil, E. J. W.; Unold, T.; Unger, E. Combinatorial Inkjet Printing for Compositional Tuning of Metal-halide Perovskite Thin Films. *J. Mater. Chem. A.* **2022**, *10*, 4906–4914.

(17) Oleneva, E.; Khaydukova, M.; Ashina, J.; Yaroshenko, I.; Jahatspanian, I.; Legin, A.; Kirsanov, D. A Simple Procedure to Assess Limit of Detection for Multisensor Synthesis. *Sensors.* **2019**, *19*, 1359.