

Note

A Study of Deep Learning for Quantitative Analysis of Vitamin A in Cattle Blood

Mizuki SHIBASAKI ^a, Tetsuhito SUZUKI ^{ab}, Moriyuki FUKUSHIMA ^a, Shin-ichi NAGAOKA ^{acd*},
Yuichi OGAWA ^{ae}, Naoshi KONDO ^a

^a Laboratory of Bio-Sensing Engineering, Division of Environmental Science and Technology, Graduate School of Agriculture, Kyoto University, Kitashirakawa-Oiwake-cho, Sakyo-ku, Kyoto 606-8502, Japan

^b Productive Environment System Laboratory, Department of Environmental Science and Technology, Graduate School of Bioresources, Mie University, 1577 Kurimamachiya-cho, Tsu, Mie 514-8507, Japan

^c Affinity Science Corporation, 1-11-1 Nishigotanda, Shinagawa-ku, Tokyo 141-0031, Japan

^d Department of Chemistry, Faculty of Science and Graduate School of Science and Engineering, Ehime University, Matsuyama, Ehime 790-8577, Japan

^e Graduate School of Agricultural Science, Tohoku University, 468-1, Aoba, Aramaki-aza, Aoba-ku, Sendai, Miyagi 980-8572, Japan

*e-mail: nagaoka.shinichi.3c@kyoto-u.ac.jp

(Received: May 4, 2024; Accepted for publication: October 28, 2024; Online publication: January 18, 2025)

Deep learning in combination with fluorescence excitation-emission spectroscopy was studied to quantitatively analyze vitamin A (retinol) in cattle blood. The neural network model being obtained with the deep learning predicted the vitamin-A levels with a coefficient of determination (R^2) of 0.93 with respect to the experimental values. The combination of the deep learning and fluorescence excitation-emission spectroscopy has a potential to predict the vitamin-A level in the cattle blood accurately, rapidly and inexpensively and to improve production of marbled beef with maintaining cattle health. It could also be applied to quantitative vitamin-A assays of various biological tissues, foods and so on as well as to those of blood samples besides cattle.

Keywords: Deep learning, Neural network, Fluorescence excitation-emission spectroscopy, Vitamin A, Retinol, Cattle blood, random forest

1 INTRODUCTION

Vitamin A (retinol, **Figure 1**) is a natural retinoid found in various biological tissues and foods, and plays a key role in vision and growth [1, 2]. Vitamin A is also linked to development, maintenance and/or function of most of organ systems in a body, and imparts color to our food. Although a variety of analytical techniques were used for vitamin A in various foods, blood samples and so on [3], high-performance liquid chromatographic (HPLC) technique has proven to be by far the method of choice [4].

Regulation of vitamin-A level in blood of Japanese black cattle “Wagyu” plays an important role in maintaining the status of “gourmet food” [5]. Meat of the Japanese black cattle is famous for presenting the highest degree of marbling, and they are fed with high-starch diets with vitamin-A deprivation to achieve the desired marbling. However, low or deprived vitamin-A diet induces

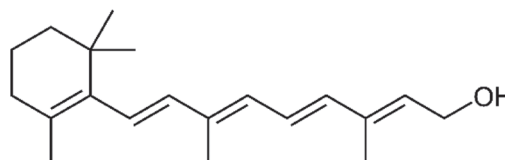


Figure 1. Molecular structure of vitamin A (retinol).

multiple negative outcomes such as occurrence of blindness and muscular edema, severe hepatic disease and swelling. Accordingly, while maintaining cattle health, a minimal blood vitamin-A level must be kept to ensure production of the marbled beef, and so Wagyu farmers monitor and regulate the level by outsourcing HPLC analysis of the blood sometimes during the fattening period.

However, the outsourcing of the conventional HPLC analysis



Open Access This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial No Derivatives (CC BY-NC-ND) 4.0 License. <http://creativecommons.org/licenses/by-nc-nd/4.0/>

is expensive and time-consuming, which makes regular frequent analysis and precise cattle-health management very difficult; the outsourcing costs about 4000 yen per one sample, and it usually takes one week. As a result, the Wagyu farmers are all eagerly looking forward to a new, accurate, rapid and inexpensive analysis-method of vitamin-A in the cattle blood, but such a convenient method is currently absent.

Accordingly, we studied fluorescence excitation-emission spectra of cattle whole blood for the quantitative vitamin-A assay, and found that fluorescence properties of vitamin A in the cattle blood were affected by internal quenching effects caused by coexisting species [6, 7]. As a result, we could perform no reliable quantitative-analysis by observing the fluorescence intensity only at a single pair of the fluorescence excitation and emission wavelengths, and we needed to use some multivariate analyses for the vitamin-A assay. On the other hand, application of deep learning with neural network to chemistry is a topic of current interest [8, 9], and its rapid spreading is remarkable [10, 11]. Therefore, to quantitatively analyze the vitamin-A we have here studied the deep learning in combination with the fluorescence excitation-emission spectroscopy, and propose a new, accurate, rapid and inexpensive method to improve the production of the marbled beef with maintaining the cattle health. Such a method could also be applied to quantitative vitamin-A assays of various biological tissues, foods and so on as well as to those of blood samples besides cattle.

2 MATERIALS AND METHODS

Sample cattle used in our study were described in detail elsewhere [6, 7] together with our experimental methods. Briefly, a total of 152 Japanese black Wagyu cattle at Tajima Agricultural High School (Yabu, Japan) were evaluated for the blood vitamin-A level during their fattening period (7–32 months). This study was carried out in strict compliance with the regulation of animal experiments at Kyoto University stated in the Guide for the Care and Use of Laboratory Animals. The regulation was approved by the Kyoto University Animal Experimentation Committee (Permit Number: R2-60). All efforts were made to minimize cattle suffering.

The cattle whole-blood sample was collected via jugular venipuncture. Within 15 min after the blood collection, the surface fluorescence emitted from the blood sample was obtained with a spectrofluorometer (JASCO (Hachioji, Japan), FP-8300) [6, 7], and an excitation-emission matrix (EEM) [12] was produced. It took about five minutes to obtain the EEM. In the EEM, a fluorescence-intensity contour-map was shown as a function of the excitation wavelength (vertical axis) and the emission wavelength (horizontal axis). Since the spectrofluorometer, together with the software given below, would be shared among agricultural cooperative members after industrialization, the financial burden on each farmer would not be so heavy. In order to extract the EEM caused by only vitamin A, from every observed EEM we subtracted the EEM that had been obtained from the whole-blood sample containing the lowest concentration of vitamin A. Details of this background subtraction are given in supplementary material. The experimental vitamin-A level in the cattle whole-blood was obtained by outsourcing the HPLC analysis to Wadayama Service Center of Hoken Kagaku

(Asago, Japan).

By using OriginPro 2024b (OriginLab (Northampton, USA)) [13], every extracted EEM was fitted with a two-dimensional Gaussian function given below.

$$z = z_0 + A \exp \left\{ -\frac{1}{2} \left(\frac{x - x_c}{w_1} \right)^2 - \frac{1}{2} \left(\frac{y - y_c}{w_2} \right)^2 \right\} \quad (1)$$

where x and y denote the fluorescence emission and excitation wavelengths, respectively. z stands for the fluorescence intensity at (x, y) in the EEM. A corresponds to extracted fluorescence peak-intensity. The emission and excitation wavelengths of the fluorescence peak are represented by x_c and y_c , and the spectral widths along the horizontal and vertical axes are represented by w_1 and w_2 , respectively. z_0 refers to background intensity.

Our deep learning analysis was performed with the holdout method [14] using Multi-Sigma Ver. 1 (Aizoth (Tsukuba, Japan)) [15]. To clarify key factors influencing prediction of the vitamin-A level, sensitivity analysis by means of the partial derivative method [16] was made using ensemble predictive models built with Multi-Sigma. Multi-Sigma performs in-cloud calculation without using a high-performance computer and without a need of programming. Because Multi-Sigma can achieve data analysis on a dataset as small as 20 and can improve accuracy of the prediction even with a small amount of sample data, the number of cattle samples could be reduced to 152 in our deep learning analysis. This reduction was in accordance with the “The Three Rs” guiding principle of replacement, reduction and refinement [17] for appropriately conducting animal experiments.

In the current study, the 152 cattle samples were divided into two groups. One group consisted of 114 samples and was used for training and validation. 12 samples randomly selected from the group were used for the validation, and the remaining 102 samples were used for the training. The second group consisted of 38 samples and was used for an independent test. The five feature values obtained from the EEM of the cattle blood (A , x_c , w_1 , y_c and w_2 in equation 1) constituted the explanatory variables, and the vitamin-A level was the objective variable. In Multi-Sigma, preprocessing configuration and predictive models were selected so as to maximize the prediction accuracy with maintaining an auto setting for the artificial intelligence configuration. It took a few hours to obtain the prediction with Multi-Sigma.

For comparison, a similar regression analysis for the prediction was performed with random forest algorithm [18] using Python and scikit-learn [19]. Random forest is an ensemble learning method and consists of multiple tree-like models of decisions (decision trees [14]) just as a forest has a lot of trees. Our random-forest analysis was performed at the default setting. Variable importance in partial least-squares (PLS) regression was evaluated by using OriginPro 2024b [13]. In principle, the model fitting with Multi-Sigma is more efficient than those with PLS and random forest, because Multi-Sigma uses a neural network algorithm. Furthermore, since Multi-Sigma tunes hyper-parameters with confirming the prediction accuracy for unknown validation data, overfitting to the training data is suppressed even in the prediction using a small

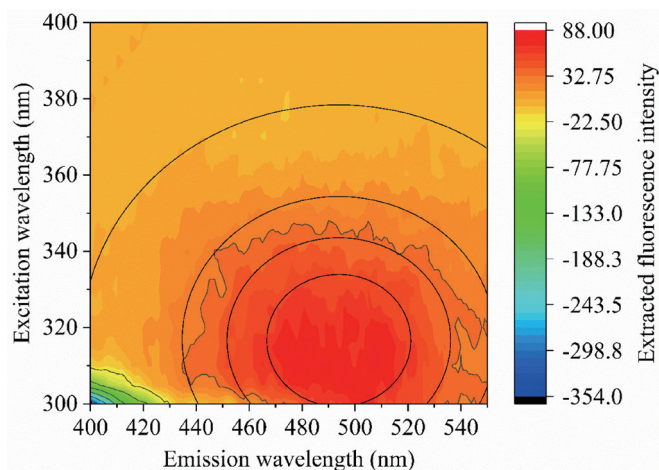


Figure 2. An example of EEM of cattle blood together with two-dimensional Gaussian-fitting result. The plot gives a two-dimensional Gaussian fit with A of 82 ± 2 , x_c of 493.9 ± 0.6 nm, w_1 of 36.8 ± 0.8 nm, y_c of 316.7 ± 0.6 nm, w_2 of 23.4 ± 0.7 nm and z_0 of -2.6 ± 0.6 .

amount of sample data, whereas PLS and random forest are liable to overfitting.

3 RESULTS AND DISCUSSION

Figure 2 shows an example of EEM together with the result of the two-dimensional Gaussian fitting explained in the preceding section. The fluorescence peak seen in the EEM is thought to come from vitamin A contained in the cattle blood [6, 7]. In fact, the x_c and y_c values are similar to the fluorescence emission- and excitation-peak wavelengths of vitamin A (retinol) [20, 21] although the y_c value shows a small blue-shift.

Figure 3a shows plots of the vitamin-A levels predicted with Multi-Sigma as functions of the experimental level. The neural network model predicts the vitamin-A levels of the test data with a coefficient of determination (R^2) of 0.93 with respect to the experimental values. The R^2 value thus estimated for the test data is greater than the one (0.91) [6, 7] obtained with PLS regression by using MATLAB (MathWorks (Natick, USA)) [22] and PLS_Toolbox (Eigenvector (Manson, USA)) [23], which need programming in contrast to Multi-Sigma requiring no programming. **Figure 3b** shows similar plots produced with the random forest method, and the estimated R^2 value is 0.91 in the plot for the test data. In each of **Figures 3a and b**, the plot for the training and validation data is also shown and the R^2 value is greater than that in the plot for the corresponding test data.

Thus, by using Multi-Sigma, the prediction accuracy of the vitamin-A level has been improved, and for farmers the analysis has become more convenient than the one reported previously [6, 7] because of the absence of programming. The prediction accuracy obtained by using Multi-Sigma is also higher than the one by the random forest method, which again requires programming. Furthermore, the use of EEM is more convenient than the outsourcing of the conventional HPLC analysis, because the EEM measurement

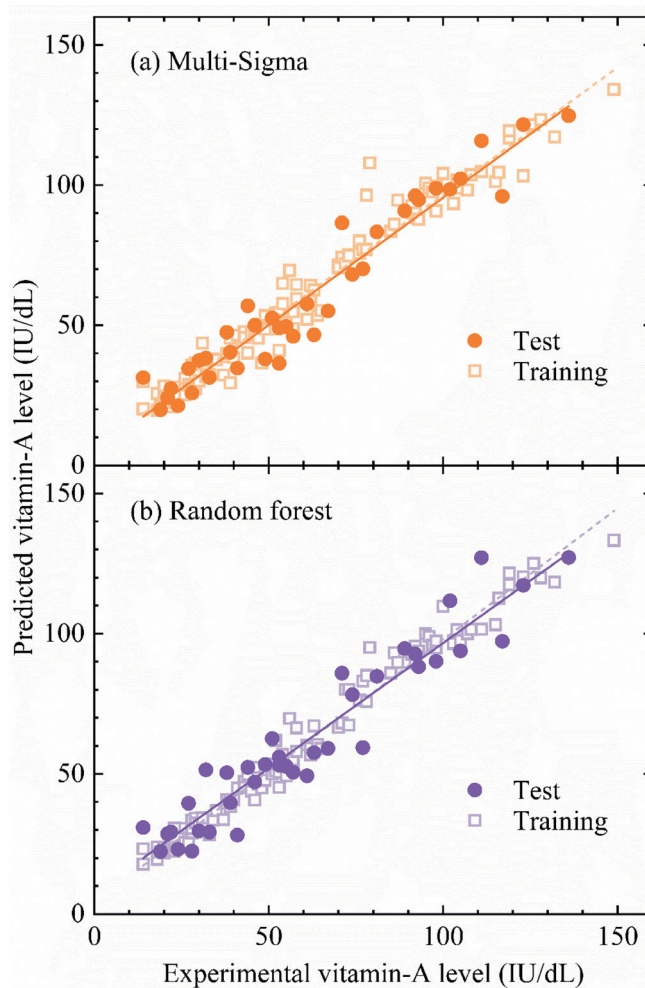


Figure 3. Plots of predicted vitamin-A levels in cattle blood as functions of the experimental level. The filled circles and the solid line with the same color denote a plot and its linear fit for the test data, respectively. The open squares and the broken line with the same color refer to those for the training and validation data. The intercepts of the linear fits may be caused by the background subtraction containing some errors. (a) Multi-Sigma. The plot for the test data (strong-orange filled circles) gives a linear fit with R^2 of 0.93, a slope of 0.91 ± 0.04 and an intercept of 5 ± 3 IU/dL (strong-orange solid line). The plot for the training and validation data (light-orange open squares) gives a linear fit with R^2 of 0.96, a slope of 0.91 ± 0.02 and an intercept of 6 ± 1 IU/dL (light-orange broken line). (b) Random forest. The plot for the test data (strong-violet filled circles) gives a linear fit with R^2 of 0.91, a slope of 0.89 ± 0.05 and an intercept of 7 ± 3 IU/dL (strong-violet solid line). The plot for the training and validation data (light-violet open squares) gives a linear fit with R^2 of 0.98, a slope of 0.94 ± 0.01 and an intercept of 4 ± 1 IU/dL (light-violet broken line).

is more rapid and inexpensive than the outsourcing. Therefore, the deep learning of EEM has a potential to predict the vitamin-A level in the cattle blood accurately, rapidly and inexpensively and to improve the production of the marbled beef with maintaining the

Table 1. Contributions of explanatory variables to predictions performed with Multi-Sigma (result of sensitivity analysis), PLS (variable importance) and random forest (feature importance)

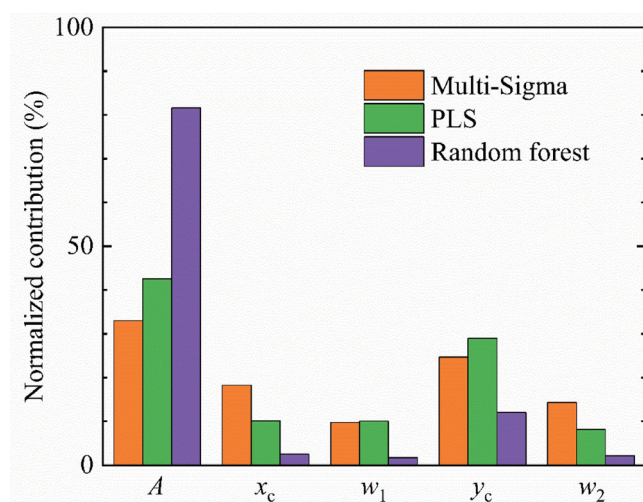
	Multi-Sigma			PLS	Random forest
	Total contribution/ %	Positive contribution/ %	Negative contribution/ %	Variable importance	Feature importance ^a
A	32.99	32.38	0.6	1.760	0.8160
x_c	18.31	17.42	0.9	0.420	0.0252
w_1	9.76	7.03	2.73	0.416	0.0171
y_c	24.66	22.24	2.42	1.200	0.1204
w_2	14.27	12.54	1.73	0.338	0.0214

^a The contribution of explanatory variables to prediction performed with random forest is called feature importance and variable importance in scikit-learn [19] and reference [18], respectively.

cattle health.

Table 1 shows contributions of the explanatory variables (A , x_c , w_1 , y_c and w_2 in equation 1) to the predictions performed with Multi-Sigma, PLS and random forest. The contribution values in Multi-Sigma [15] have been obtained from the sensitivity analysis, and those in PLS [13] and random forest [19] correspond to the variable importance and feature importance, respectively. In the results from all the prediction methods given in **Table 1**, the A value (fluorescence peak intensity) shows the greatest contribution among all the explanatory variables. In the result of the sensitivity analysis made with Multi-Sigma, the positive contribution of the A value is much greater than the negative contribution. Accordingly, the fluorescence peak intensity is thought to positively correlate with the vitamin-A level, which is consistent with our premise that the fluorescence comes from vitamin A contained in the cattle blood. Since the positive contributions of the x_c , w_1 , y_c and w_2 values are greater than the negative ones, the vitamin-A level would increase as the fluorescence emission and excitation show red-shift and the spectral widths increase, which is consistent with the results obtained previously [6, 7].

In each of the columns of PLS and random forest in **Table 1**, the listed importance values were summed, each of the importance values was divided by the summation, and the result was multiplied by 100; that is, the importance was represented in the unit of %. The value thus obtained was regarded as the normalized contribution of the corresponding explanatory variable to the prediction, together with the total contribution value estimated and normalized with Multi-Sigma (**Table 1**). **Figure 4** shows comparison among these normalized contributions obtained with Multi-Sigma, PLS and random forest. In random forest (violet bar charts), the prediction seems to be basically made on the basis of only the A value, because the normalized contribution of A is over 80%. In Multi-Sigma and PLS (orange and green bar charts, respectively), the normalized contributions of A and y_c are much greater than those of the other explanatory variables (x_c , w_1 and w_2), and the contributions of A and y_c in Multi-Sigma are less than those in PLS. In contrast to these relatively-great normalized-contributions (A and y_c), the contributions of the explanatory variables x_c , w_1 and w_2 to the predictions are relatively-small, and those in Multi-Sigma are greater than, or close to, the ones in PLS. These results may suggest that Multi-Sigma does not overlook even such relatively-small contributions.

**Figure 4.** Comparison among normalized contributions of explanatory variables to predictions performed with Multi-Sigma, PLS and random forest.

4 SUMMARY

The deep learning in combination with the EEM measurement was studied to quantitatively analyze vitamin A (retinol) in the cattle blood. The neural network model being obtained with the deep learning predicted the vitamin-A levels with an R^2 of 0.93 with respect to the experimental values. The deep learning of EEM has a potential to predict the vitamin-A level in the cattle blood accurately, rapidly and inexpensively and to improve the production of the marbled beef with maintaining the cattle health. Furthermore, this method could also be applied to quantitative vitamin-A assays of various biological tissues, foods and so on as well as to those of blood samples besides cattle.

5 SUPPLEMENTARY MATERIAL

The method and procedure for extracting the EEM intensity of only vitamin A are provided as supplementary material and available online.

ACKNOWLEDGMENTS

We express our sincere thanks to Mr. Takahiko Ohmae and Mr. Norio Nishiki of Tajima Agricultural High School for their kind helps in the evaluation of the vitamin-A level in the cattle blood. Our thanks are also due to Dr. Kotaro Kawajiri of Aizoth Inc. for providing support to our deep learning analyses. S.N. thanks Professor Yasushi Minowa of Kyoto Prefectural University for his valuable discussion, and also thanks Mr. Tomohiko Tasaka, president of Affinity Science Corp., for his continuous encouragement. This work was partly supported by JSPS KAKENHI Grant Numbers 20H00439 and 23H00350.

REFERENCES

- [1] C. P. F. Redfern, *Methods Enzymol.*, **637**, 1 (2020). <https://doi.org/10.1016/bs.mie.2020.02.002>, PMID:32359642
- [2] B. S. Gebregziabher, H. Gebremeskel, B. Debesa, D. Ayalneh, T. Mitiku, T. Wendwessen, *et al.*, *J. Agric. Food Res.*, **14**, 100834 (2023). <https://doi.org/10.1016/j.jafr.2023.100834>
- [3] E.-S. Tee, C.-L. Lim, *Food Chem.*, **41**, 147 (1991). [https://doi.org/10.1016/0308-8146\(91\)90042-M](https://doi.org/10.1016/0308-8146(91)90042-M)
- [4] P. V. Bhat, R. R. Sundaresan, *Crit. Rev. Anal. Chem.*, **20**, 197 (1988). <https://doi.org/10.1080/00078988808048812>
- [5] L. Zhao, X. Liu, N. A. Gomez, Y. Gao, J. S. Son, S. A. Chae, *et al.*, *J. Anim. Sci. Biotechnol.*, **14**, 2 (2023). <https://doi.org/10.1186/s40104-022-00805-0>, PMID:36597116
- [6] M. Shibasaki, T. Suzuki, Y. Saito, M. Fukushima, T. Fujiura, T. Ohmae, N. Nishiki, N. Kondo, *J. Jpn. Soc. Agric. Machinery Food Eng.*, **83**, 477 (2021) (in Japanese). https://doi.org/10.11357/jsamfe.83.6_477
- [7] M. Shibasaki, T. Suzuki, Y. Saito, N. Li, M. Fukushima, Y. Ogawa, *et al.*, *J. Jpn. Soc. Agric. Machinery Food Eng.*, submitted (in Japanese)
- [8] H. Teramae, T. Matsuo, K. Niwatsukino, R. Inoue, S. Noguchi, M. Xuan, T. Yamashita, J. Takayama, M. Okazaki, T. Sakamoto, *J. Comput. Chem. Jpn.*, **19**, 43 (2020) (in Japanese). <https://doi.org/10.2477/jccj.2020-0005>
- [9] H. Teramae, *J. Comput. Chem. Jpn.*, **22**, 34 (2023) (in Japanese). <https://doi.org/10.2477/jccj.2023-0022>
- [10] S. Shilpa, G. Kashyap, R. B. Sunoj, *J. Phys. Chem. A*, **127**, 8253 (2023). <https://doi.org/10.1021/acs.jpca.3c04779>, PMID:37769193
- [11] B. Debus, H. Parastar, P. Harrington, D. Kirsanov, *Trends Anal. Chem.*, **145**, 116459 (2021). <https://doi.org/10.1016/j.trac.2021.116459>
- [12] Y. Tamura, H. Inoue, S. Takemoto, K. Hirano, K. Miyaura, *J. Fluoresc.*, **31**, 91 (2021). <https://doi.org/10.1007/s10895-020-02640-w>, PMID:33094367
- [13] OriginPro, 2024b; OriginLab: Northampton, MA, USA, 2024. <https://www.originlab.com/index.aspx?go=PRODUCTS/Origin> (accessed December 23, 2024).
- [14] I. H. Witten, E. Frank, M. A. Hall, *Data Mining, Practical Machine Learning Tools and Techniques*, 3rd ed.; Elsevier: Amsterdam, 2011; (a) Chapter 5. (b) Chapter 3.
- [15] Multi-Sigma Ver, 1; Aizoth: Tsukuba, Japan, 2023. <https://aizoth.com/en/service/multi-sigma/> (accessed December 23, 2024).
- [16] M. Gevrey, I. Dimopoulos, S. Lek, *Ecol. Modell.*, **160**, 249 (2003). [https://doi.org/10.1016/S0304-3800\(02\)00257-0](https://doi.org/10.1016/S0304-3800(02)00257-0)
- [17] W. M. S. Russell, R. L. Burch, *The Principles of Humane Experimental Technique*, 1959. <https://caat.jhsph.edu/principles/the-principles-of-humane-experimental-technique>. (accessed December 23, 2024).
- [18] L. Breiman, *Mach. Learn.*, **45**, 5 (2001). <https://doi.org/10.1023/A:1010933404324>
- [19] Boritaso. Random Forest, 2024 (in Japanese). https://boritaso-blog.com/random_forest/ (accessed December 23, 2024).
- [20] S. Futterman, J. Heller, *J. Biol. Chem.*, **247**, 5168 (1972). [https://doi.org/10.1016/S0021-9258\(19\)44953-3](https://doi.org/10.1016/S0021-9258(19)44953-3), PMID:5066567
- [21] P. A. Peterson, L. Rask, *J. Biol. Chem.*, **246**, 7544 (1971). [https://doi.org/10.1016/S0021-9258\(19\)45809-2](https://doi.org/10.1016/S0021-9258(19)45809-2), PMID: 5002681
- [22] Partial least-squares (PLS) regression in MATLAB; MathWorks: Natick, MA, USA, 2024. <https://jp.mathworks.com/help/stats/plsregress.html?lang=en> (accessed December 23, 2024).
- [23] PLS_Toolbox; Engenvektor: Manson, WA, USA, 2024. https://jp.mathworks.com/products/connections/product_detail/pls-toolbox.html (accessed December 23, 2024).