Non-Destructive Classification of Sake Acidity Using Excitation–Emission Matrix Combined with Chemometrics

Pornpimol Moolkaew, Yoshito Saito*, Takumi Murai, Yu Obata, and Hideo Hasegawa

Graduate School of Science and Technology, Niigata University, Japan
Email: f23n501d@mail.cc.niigata-u.ac.jp (P.M.); ysaito@agr.niigata-u.ac.jp (Y.S.); f24e022a@mail.cc.niigata-u.ac.jp (T.M.); f24e011e@mail.cc.niigata-u.ac.jp (Y.O.); hsgw@agr.niigata-u.ac.jp (H.H.)
*Corresponding author
Manuscript received November 14, 2025; accepted December 1, 2025; published June 29, 2026

***Abstract*—Sake acidity is a critical quality parameter that significantly influences taste perception and consumer preference. Conventional titrimetric methods, while accurate, are labor-intensive, time-consuming, and destructive, limiting their suitability for rapid quality assessment. This study demonstrates the feasibility of Excitation–Emission Matrix (EEM) fluorescence spectroscopy combined with chemometric modeling for rapid, non-destructive classification of sake acidity. A total of 100 commercial sake samples were analyzed and classified into low- and high-acidity groups using Partial Least Squares Discriminant Analysis (PLS-DA) and principal component analysis–support vector machine (PCA-SVM). The optimized PLS-DA model with seven latent variables achieved superior classification performance (81% accuracy, F1-scores >80%), substantially outperforming PCA-SVM variants (58–75% accuracy). VARIABLE Importance in Projection (VIP) analysis revealed that aromatic amino acid fluorescence (Peak A, Ex/Em 265–305/285–420 nm) and Maillard reaction products (Peak B, Ex/Em 315–350/390–445 nm) served as the primary discriminators, with riboflavin regions (Peak C, Ex/Em 300–355/470–580 nm) contributing secondarily. High-acidity sake exhibited enhanced Maillard-derived fluorescence, whereas low-acidity sake retained stronger aromatic amino acid signals. This fluorescence-based approach provides a rapid, reagent-free alternative to conventional titration, offering practical potential for real-time brewery quality control and production monitoring without sample consumption.**

***Keywords*—Chemometrics, EEM fluorescence spectroscopy, PLS-DA, Sake acidity**

I. Introduction

The major constituents of sake are water, ethanol (generally 13–17%), carbohydrates, and amino acids, along with a small proportion of organic acids. Although organic acids account for only about 0.5% of the total composition, with typical acidity values ranging from 1.0 to 2.0, they are primarily produced by yeast during the main mash fermentation and most of these are composed of succinic, lactic, malic, and acetic acids. These compounds play a crucial role in shaping the flavor, aroma, and color of sake, and also act as key factors in inhibiting microbial growth [1]. Sake with high acidity is typically perceived as sharp and dry, as sourness can effectively suppress the perception of sweetness. Conversely, reducing the acidity enhances sweetness, giving low-acidity sake a richer and smoother taste [2]. Although acidity is widely recognized as an important indicator for flavor profiling and product differentiation, no standardized criterion currently defines the threshold between high- and low-acidity sake. Nonetheless, acidity-based categorization remains practically significant for breweries, researchers, and quality control purposes, as it supports process optimization, sensory evaluation, and market-oriented classification. Therefore, developing a reliable method to distinguish sake according to its acidity level continues to be a subject of scientific and industrial interest.

Conventional methods for assessing sake acidity rely primarily on titration and chromatographic techniques such as high-performance liquid chromatography [2]. Although titration is simple, inexpensive, and widely used in breweries, it measures only the total free organic acids and cannot distinguish individual compounds. Chromatographic analysis, on the other hand, provides molecular-level identification and quantification of specific organic acids; however, it is time-consuming, requires labor-intensive sample preparation and chemical reagents, generates laboratory waste, and is unsuitable for rapid or routine measurements. In contrast, non-destructive spectroscopic techniques offer rapid measurement, minimal sample preparation, and greater environmental sustainability, with the potential for real-time quality monitoring. Methods such as Near-Infrared (NIR) and Fourier-Transform Infrared (FT-IR) spectroscopy have been successfully applied to evaluate pH, total acidity, and organic acids in alcoholic beverages such as wine [3, 4].

Nonetheless, both NIR and FT-IR are often limited by overlapping absorption bands in complex matrices such as fermented beverages, which can reduce their selectivity for minor organic acids. Fluorescence-based techniques, particularly Excitation–Emission Matrix (EEM) spectroscopy combined with chemometrics, have therefore gained attention as a more sensitive and selective alternative for acidity characterization in alcoholic products. EEM fluorescence spectroscopy has been widely applied in wine analysis, where its molecular fingerprints from phenolics, pigments, vitamins, and aromatic amino acids combined with chemometrics enable accurate variety and geographical classification [5, 6]. The primary organic acids present in sake, including succinic, lactic, malic, and acetic acids, are non-fluorescent under typical excitation conditions. However, variations in their concentrations alter the overall acidity of the matrix, thereby modulating the protonation state and microenvironment of naturally fluorescent compounds such as amino acids, phenolic acids, and other aromatic fluorophores. Consequently, changes in organic acid composition are indirectly reflected in the EEM spectral profile through pH-dependent quenching and matrix effects. Because EEM spectra are inherently high-dimensional and contain strongly overlapped fluorescence signals, direct visual interpretation is insufficient for extracting meaningful chemical information. Chemometric approaches are therefore

required to decode these complex fingerprints by reducing dimensionality, suppressing noise, and highlighting subtle spectral differences. Principal component analysis (PCA) is widely applied for exploratory visualization and feature reduction, while supervised classification techniques such as support vector machines (SVM) and partial least squares discriminant analysis (PLS-DA) have demonstrated strong performance in food authenticity, quality classification, and contamination detection. For instance, PCA-SVM accurately discriminated male-fertile and male-sterile *Cryptomeria japonica*, while EEM-based PLS-DA successfully identified authentic and adulterated honey samples [7, 8].

The objectives in this study is non-destructive classification of sake acidity by integrating excitation–emission matrix (EEM) fluorescence spectroscopy with chemometric modeling. The specific approaches are threefold: (1) to develop and compare classification models based on PLS-DA and PCA-SVM for distinguishing high- and low-acidity sake samples; (2) to identify the key fluorescence features responsible for acidity discrimination through variable importance analysis; and (3) to evaluate the practical feasibility of this fluorescence-based approach as a rapid, reagent-free alternative to conventional titrimetric methods for brewery quality control.

II. Materials and Methods

A. Preparation of Sake Samples

A total of 100 commercial sake samples, originating from different production regions of Japan, were purchased in four batches from online retailers. The samples were obtained in different types of packaging, including cans, plastic bottles, and glass bottles (amber, green, and clear) with various volumes. All samples were stored unopened in a dark environment under ambient room temperature conditions to avoid light-induced degradation until analysis. Before measurement, each sample was gently mixed and analyzed without any further pretreatment.

B. Acidity Analysis of Sake

The titratable acidity of the sake samples was determined following the method described in [9]. A 10 mL aliquot of each sample was titrated in duplicate with standardized 0.1 M NaOH to an endpoint of pH 7.2, measured using a calibrated pH meter (ST20, OHAUS, USA). Titratable acidity reflects the concentration of free organic acids, such as malic, lactic, and succinic acids, which contribute to the overall acidity of sake. The titratable acidity was recorded and expressed as the mean volume (mL) of 0.1 M NaOH required per 10 mL of sample, in accordance with [9].

C. Collection of EEM Fluorescence Spectra

1) EEM fluorescence acquisition

EEM fluorescence spectra of the sake samples were acquired using a spectrofluorometer (FP-8550, JASCO Corporation, Japan) operated in a right-angle configuration, which is commonly employed for transparent liquid samples to minimize scattering effects. A 3 mL aliquot of each sample was transferred into a standard quartz cell (Q-204, ASLAB Co., Ltd., Japan) prior to measurement. Each sample was scanned in triplicate over an excitation (Ex) range of 210–535 nm and an emission (Em) range of 250–750 nm, with both the wavelength interval and the spectral bandwidth set to 5 nm.

EEM spectra were corrected using reference measurements from a halogen light source and rhodamine B to account for instrumental bias. In addition, the Raman scattering of distilled water (Ex = 350 nm, Em = 371–428 nm) was recorded each day and used to normalize the data, minimizing variations in detector sensitivity among measurement sessions [10].

2) Pretreatment of EEM

Raw EEM data were preprocessed prior to chemometric modeling to suppress non-fluorescent artifacts and stabilize distributional properties of the intensities. All analyses were performed in MATLAB R2024a (MathWorks, USA). Regions affected by first- and second-order Rayleigh scattering, as well as stray light at short Ex wavelengths, were removed by hard masking (set-to-zero) based on their respective Ex–Em positions. This step prevents strong scatter bands from dominating the fluorescence landscape and ensures that only meaningful spectral information is retained. The preprocessed three-dimensional EEMs were then unfolded into two-dimensional sample-wise row vectors to form the input matrix for subsequent chemometric classification.

3) Chemometric Classification Models

The acidity values of the 100 sake samples were divided into two classes, Low and High, using the median as the threshold. Samples with acidity values above the median were assigned to the High-acidity group, whereas those below the median were assigned to the Low-acidity group. In this study, class labels were numerically encoded as High = 1 and Low = 0 to facilitate model computation and interpretation. Two chemometric classification approaches were employed in this study: PLS-DA and PCA-SVM.

PLS-DA is a supervised classification method that projects the predictor matrix (unfolded EEM fluorescence intensity data, X) into a set of latent variables (LVs) that maximizes the covariance between the X and the categorical response variable (acidity level, Y), PLS-DA considers the class labels during dimension reduction, thereby enhancing class separation in the latent space. The optimal number of LVs was determined through k-fold cross-validation by selecting the model that yielded the minimum root mean square error of cross-validation (RMSECV), balancing model complexity and predictive performance [11] A classification threshold of 0.5 was used to distinguish the two categories during model prediction. Variable importance in projection (VIP) scores were calculated to identify the most influential spectral regions, where VIP values greater than 1 indicate above-average importance.

For the PCA-SVM workflow, PCA was applied to reduce the dimensionality of the unfolded EEM data by extracting the most informative features while eliminating noise and redundancy. This unsupervised method transforms the original correlated fluorescence variables into a new set of orthogonal principal components (PCs) that capture the maximum variance within the data [12]. The resulting PCs scores were subsequently used as input variables for SVM to classify the sake samples according to their acidity levels. SVM constructs an optimal separating hyperplane that

maximizes the margin between classes in a high-dimensional feature space, and a linear kernel was employed in this study to generate the decision boundary. A 10-fold cross-validation was applied to both the PLS-DA and SVM models.

4) Model evaluation criteria

Model performance was evaluated using accuracy, precision, recall, and F1-score. Accuracy reflects the overall correctness of classification, precision measures the reliability of positive predictions, and recall evaluates the model's ability to detect true positives. The F1-score balances both precision and recall. These evaluation parameters are calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - score = \frac{2\ x\ Precision\ x\ Sensitivity}{Precision\ + Sensitivity} \quad (4)$$

The 4-evaluation metrics are calculated based on the confusion matrix, using the following definitions: TP (true positive) refers to positive samples that are correctly classified as positive; FP (false positive) refers to negative samples that are incorrectly classified as positive; TN (true negative) represents negative samples that are correctly classified as negative; and FN (false negative) refers to positive samples that are misclassified as negative. These parameters collectively provide a comprehensive assessment of the model's classification capability.

III. Results and Discussion

A. Acidity of Sake

The titratable acidity of 100 commercial sake samples ranged from 1.0 to 2.9 (mean = 1.61 ± 0.36), with alcohol content ranging from 14–20% (v/v), which aligns with the typical acidity range (1.3–2.7) reported in previous studies [2]. Compared with aged sake samples reported by [9], which exhibited a broader acidity range of 1.1–6.1, the narrower rang observed in this study suggests that the sakes analyzed were relatively less influenced by long-term storage or oxidative reactions that increase organic acid content. The acidic character in sake is attributed to the presence of weak organic acids, primarily lactic, succinic, malic, and citric acids, produced during fermentation by *Saccharomyces cerevisiae*. Previous studies have consistently reported lactic acid as the dominant organic acid in Japanese sake, typically ranging from 510–1216 mg/L [2].

Table 1. Characteristic fluorescence peaks and their corresponding compounds in sake samples.

Peak	Ex (nm)	Em (nm)	Attributes	Ref.
Peak A	265–305	285–420	Aromatic amino acids such as tyrosine, phenylalanine, and tryptophan	[13]
Peak B	315–350	390–445	Maillard reaction products	[14]
Peak C	300–355	470–580	Riboflavin (vitamin B2)	[15]
Peak D	280–315	400–465	Fulvic-like and humic-like	[16]
Peak E	340–380	405–475	Maillard reaction products	[14]

The differences in acidity may reflect variations in fermentation conditions, such as the rice-to-koji ratio and the type of koji used, which are known to affect the synthesis of specific organic acids [17]. As acidity increases, the perceived dryness of sake intensifies, while lower acidity enhances sweetness perception [2]. However, total acidity does not directly correspond to the active acidity (pH), as it encompasses both dissociated and undissociated forms of these acids. The measured values thus represent the titratable capacity of the sake matrix rather than the immediate hydrogen ion concentration. To enable binary classification, the samples were divided into two categories based on the median acidity value (1.55 mL). Samples with acidity values above this threshold were assigned to the high-acidity group (n = 49), whereas those below the median were classified as the low-acidity group (n = 51)

B. Fluorescence characteristics of sake samples

The EEM fluorescence spectra of 100 sake samples were measured to investigate their intrinsic fluorophore characteristics. The representative results are illustrated in Fig. 1, which displays the Ex/Em wavelength and fluorescence intensity distributions. According to the spectral profiles, the samples were categorized into three fluorescence patterns: Type I, Type II, and Type III. Among the total samples, 70 exhibited Type I characteristics, 8 belonged to Type II, and 22 were classified as Type III. Each fluorescence peak represents the dominant fluorophore compounds present in the sake samples, as summarized in Table 1. It should be noted that these patterns represent the typical fluorescence profiles; however, slight spectral variations were also observed within each type. For example, some Type I samples exhibited an additional emission band corresponding to Peak E.

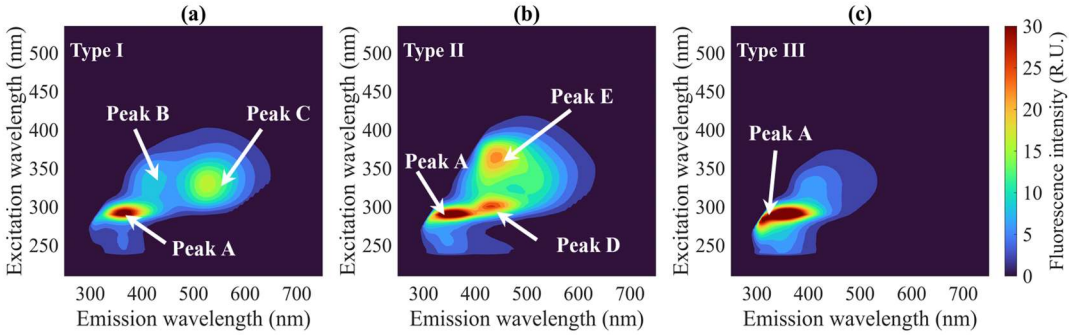


Fig. 1. Three-dimensional EEM fluorescence spectra of samples classified as (a) Type I, (b) Type II, and (c) Type III. The color scale represents fluorescence intensity in relative units (R.U.).

Peak A, which corresponds to amino acids, was detected in all sake types and is mainly derived from the *koji*-making process, where enzymes produced by *Aspergillus oryzae* hydrolyze rice proteins into free amino acids [18]. In some sake samples, a fluorescence band corresponding to humic-like substances (Peak D) was also detected. This signal may be attributed to trace humic compounds introduced through the water used for dilution (*warimizu*), as reported previously [19]. Humic acids naturally present in water can enter sake during production and act as photosensitizers, promoting color development and humic-like fluorescence during storage [19].

The absence of riboflavin in certain samples may result from photodegradation, removal during fining processes with bentonite or activated carbon, or strain-dependent differences in yeast riboflavin production. In addition, packaging transparency can accelerate riboflavin degradation under light exposure [20, 21], which may further reduce its detectable fluorescence in sake.

C. Classification of Sake Acidity by PLS-DA

Table 2 summarizes the classification performance of PCA-SVM and PLS-DA models for distinguishing high- and low-acidity sake samples. The optimized PLS-DA model, constructed with seven latent variables (LV=7), achieved robust classification performance with 81% overall accuracy. Both acidity classes exhibited balanced discrimination: high-acidity samples showed 80% precision and 81.63% recall (F1-score: 80.81%), while low-acidity samples exhibited 82% precision and 80.39% recall (F1-score: 81.19%). The nearly identical metrics confirm that the model is free from class bias and suitable for practical quality-control applications.

VIP analysis (Fig. 2a) identified the specific spectroscopic features responsible for acidity discrimination. Regions with the highest VIP scores (VIP > 2, red zones) were primarily localized to protein-like fluorescence areas (tryptophan, tyrosine) and Maillard reaction product regions (Peak B), indicating their dominant role in distinguishing acidity levels. Riboflavin regions exhibited moderate VIP scores (VIP = 1–2, yellow zones), contributing secondarily to classification. The first latent variable (LV1, 57.36%) captured an inverse relationship between acidity and fluorescence intensity in the amino acid region, indicating that lower-acidity samples retain higher concentrations of free amino acids. One of the reason of the inverse relationship can be explained by quenching through organic acid anions ($RCOO^-$), which form ion-pair complexes with protonated tryptophan at our sake pH (3.95–4.70), exhibiting both static and dynamic quenching. The quenching mechanisms may involve electrostatic perturbation of the indole excited state, enhanced non-radiative decay, and possibly weak excited-state interactions, though the contribution of electron transfer cannot be definitively established without further investigation [22].

Table 2. Classification performance of PCA-SVM and PLS-DA models.

Model	Input set	Class	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
PCA-SVM	PC1 and PC2	High	75.00	74.00	75.51	74.75
		Low		76.00	74.51	75.25
PCA-SVM	PC1 and PC3	High	74.00	72.55	75.51	74.00
		Low		75.51	72.55	74.00
PCA-SVM	PC2 and PC3	High	58.00	57.78	53.06	55.32
		Low		58.18	62.75	60.38
PCA-SVM	PC1 and PC2 and PC3	High	74.00	72.55	75.51	74.00
		Low		75.51	72.55	74.00
PLS-DA	LV=7	High	81.00	80.00	81.63	80.81
		Low		82.00	80.39	81.19

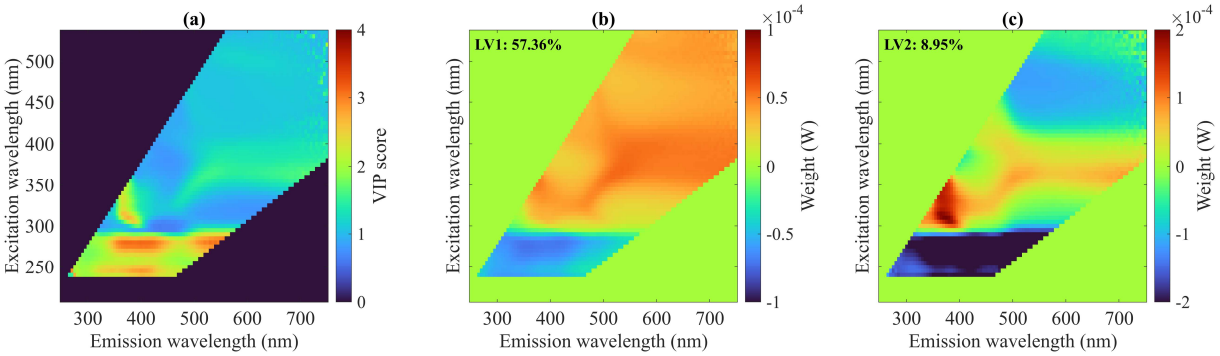


Fig. 2. Spectral feature importance analysis from PLS-DA classification model: (a) VIP score distribution across the EEM, with higher scores (red regions) indicating greater contribution to acidity discrimination; (b, c) loading weight maps for latent variables 1 and 2, explaining 57.36% and 8.95% of spectral variance, respectively. Red and blue regions represent positive and negative weights contributions to class separation, respectively.

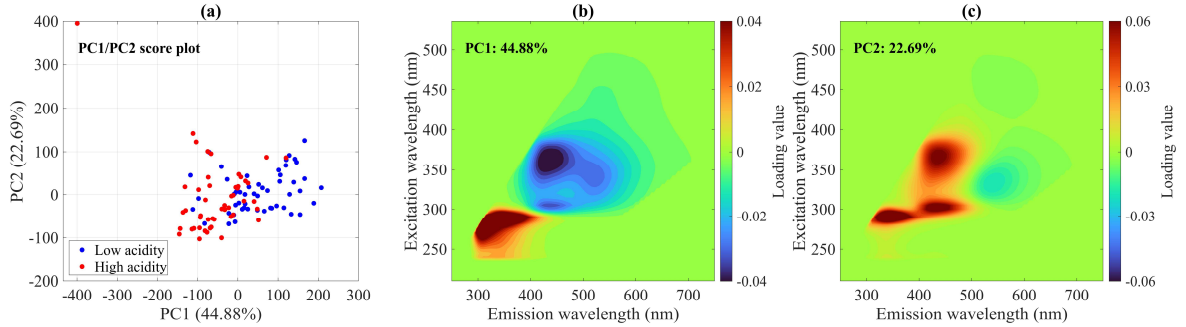


Fig. 3. PCA of EEM fluorescence spectra from sake samples: (a) PC1/PC2 score plot showing partial separation between low- and high-acidity groups; (b) PC1 loading spectrum (44.88% explained variance); (c) PC2 loading spectrum (22.69% explained variance). Red and blue color in loading plots indicate positive and negative contributions, respectively.

Conversely, the Maillard region showed a positive relationship: higher-acidity samples displayed elevated Maillard-derived fluorescence. This reflects concomitant formation of organic acids (acetic acid) and Maillard products during saccharide degradation [23]. Reactive intermediates such as methylglyoxal drive protein aggregation and cross-linking, generating fluorescent melanoidins and advanced glycation end products (AGEs) that accumulate with Maillard reaction intensity. Thus, higher organic acid levels indirectly indicate enhanced browning and accumulation of Maillard fluorophores. The opposite trend between protein-like and Maillard regions may be attributed to protein aggregation and fluorescence quenching of quenching of aromatic residues, particularly tryptophan and tyrosine.

The PLS-DA loadings (Fig. 2b, LV1) further indicate that samples with higher acidity display stronger Peak C (riboflavin) intensity. This positive association is consistent with the co-occurrence of riboflavin and fermentation-derived organic acids, both produced through active yeast metabolism. Nevertheless, the observed relationship most likely reflects their shared metabolic origin rather than a direct biochemical linkage between riboflavin and acidity.

D. Classification of Sake Acidity by PCA-SVM

PCA was applied to explore spectral variations between high- and low-acidity sake samples based on EEM fluorescence data. The first three PCs explained 86.38% of the total variance (PC1: 44.88%, PC2: 22.69%, PC3: 18.81%). The PC1–PC2 score plot (Fig. 3a) shows partial separation, with low-acidity samples clustering at positive PC1 scores and high-acidity samples at negative ones. Some overlap remains, indicating that factors other than acidity also contribute to spectral variability, while PC2 offers limited discrimination.

The loading plot for PC1 (Fig. 3b) reveals two distinct spectral regions with contrasting contributions to the PCs. The region characterized by Ex/Em 265–305/285–420 nm

displays strong positive loadings (red region, loading value ≈ +0.04). This spectral signature is consistent with protein-like fluorescence. Conversely, the region at Ex/Em 340–380/405–475 nm exhibits notable negative loadings (blue

region, loading value ≈ -0.04). This spectral feature is characteristic of Maillard reaction products, which are formed through non-enzymatic browning reactions between amino acids and reducing sugars. The subsequent negative loadings were primarily linked to fulvic-like and humic-like substances, together with minor riboflavin-related fluorescence.

Integration of score and loading plots reveals clear compositional differences between the two acidity groups. High-acidity samples (negative PC1) show stronger fluorescence in the Maillard reaction region, indicating higher levels of browning products. In contrast, low-acidity samples (positive PC1) exhibit enhanced protein-like fluorescence, reflecting greater free amino acid content. These trends align with the PLS-DA findings.

Humic-like fluorescence (Peak D) showed opposite loading directions between PC1 and PC2, reflecting the intrinsic structural diversity of humic-like substances. These compounds, derived from biopolymers such as lignin and cellulose, contain various aromatic, polycyclic, hydroxyl, and carbonyl functional groups [24], resulting in a heterogeneous ensemble of fluorophores with overlapping emission characteristics. In PC1, humic-like signals co-occurred with

Maillard products, indicating that increased humic-like fluorescence under acidic conditions may contribute to the same spectral trend observed for high-acidity samples. Conversely, PC2 revealed an inverse correlation with acidity, where lower-acidity samples exhibited stronger humic-like signals, consistent with the buffering contribution of water-derived humic substances [19]. These contrasting trends likely reflect the coexistence of humic-like fluorophores different molecular structures within the broad Ex/Em region. Given that the PLS-DA model did not prioritize this feature and that it appeared in only ~6% of the total samples, its overall impact on acidity classification appears minimal.

The PCA-SVM classification results (Table 2) demonstrate that PC1 is the primary discriminant axis for acidity level, with models incorporating PC1 achieving substantially higher accuracy (74-75%) compared to the PC2 and PC3 model (58%). This confirms that the biochemical transformations captured by PC1 specifically, the inverse relationship between amino acid fluorescence and Maillard reaction products (Fig. 3b) are the dominant features distinguishing high- and low-acidity sake samples. The addition of PC2 to PC1 improved classification accuracy marginally, this indicates that PC2 captures secondary spectral variations such as humic-like fluorescence and tryptophan region (Fig. 3c), provides supplementary but nonessential discriminatory information. Notably, the inclusion of PC3 did not enhance performance (PC1 and PC2 and PC3: 74%), indicating that higher-orde r components likely represent noise or sample-specific variations unrelated to acidity.

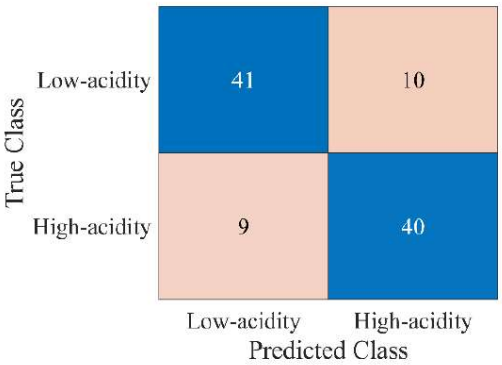
PLS-DA achieved the highest classification performance, outperforming all PCA-SVM models. The superior accuracy of PLS-DA arises from its supervised extraction of latent variables that maximize covariance with the class labels, effectively capturing acidity-related spectral variance (Fig. 2). In contrast, PCA-SVM relies on unsupervised variance maximization and a linear kernel, which may not fully represent the nonlinear relationships within the fluorescence data. The optimal PLS-DA model with LV = 7 (LV1 = 57.36%, LV2 = 8.95%) therefore provided a more chemically meaningful separation than the variance-based PCs (PC1–PC3 = 44.88–18.81%).

E. Spectral Characteristics of Misclassified Samples

To elucidate the origins of classification errors in the PLS-DA model, the average EEM fluorescence spectra of all 19 misclassified samples were analyzed according to their EEM spectrum type and misclassification direction (Figs. 4-5). The results reveal that the majority of misclassification patterns could be rationalized in terms of biochemical heterogeneity and the variable importance distribution obtained from the VIP analysis (Fig. 2a).

Type I spectra, characterized by Peaks A (aromatic amino acids), B (Maillard products), and C (riboflavin), accounted for the majority of misclassified samples (12/19, 63%). False positives (n=7, Fig. 5a) exhibited reduced Peak A combined with elevated Peaks B-C, corresponding to low-acidity sakes with strong non-enzymatic browning or riboflavin accumulation. Conversely, false negatives (n=5, Fig. 5b) displayed elevated Peak A with reduced Peaks B-C, indicating incomplete quenching despite high acidity. These results confirm that Peak A intensity reflects amino acid concentration and protonation, whereas Peaks B–C represent Maillard- and riboflavin-related reactions governed by fermentation chemistry. Type II misclassifications were associated with fluorescence features located in Peaks A, D, and E. Although Peaks D–E (humic-like and secondary Maillard fluorophores) had very low VIP scores, they still influenced predictions when combined with low Peak A intensity. Type II spectra exhibited the highest misclassification rate (37.5%, 3/8 samples), primarily due to the limited number of training samples (n=8, 8% of dataset) and the inherently ambiguous origin of humic-like fluorescence, suggesting potential interference from humic-like substances introduced during sake production, possibly derived from the water used in brewing [19].

(a)



Misclassification Type	EEM Type I	EEM Type II	EEM Type III
Fale Positive (n=10) (True Low→ Predicted High)	7	2	1
False Negative (n=9) (True High→ Predicted Low)	5	1	3
Total misclassified	12	3	4
*Misclassification rate (%)	17%	37.5%	18%

* Denominator: Total samples with that EEM type (Type I: 70, Type II: 8, Type III: 22)

(b)

Fig. 4. Classification performance and error analysis of the optimized PLS-DA model: (a) confusion matrix showing correct classifications and misclassifications for low- and high-acidity sake (n = 100, 81% accuracy); (b) distribution of EEM spectral patterns among the 19 misclassified samples, categorized into three fluorescence types with misclassification rates.

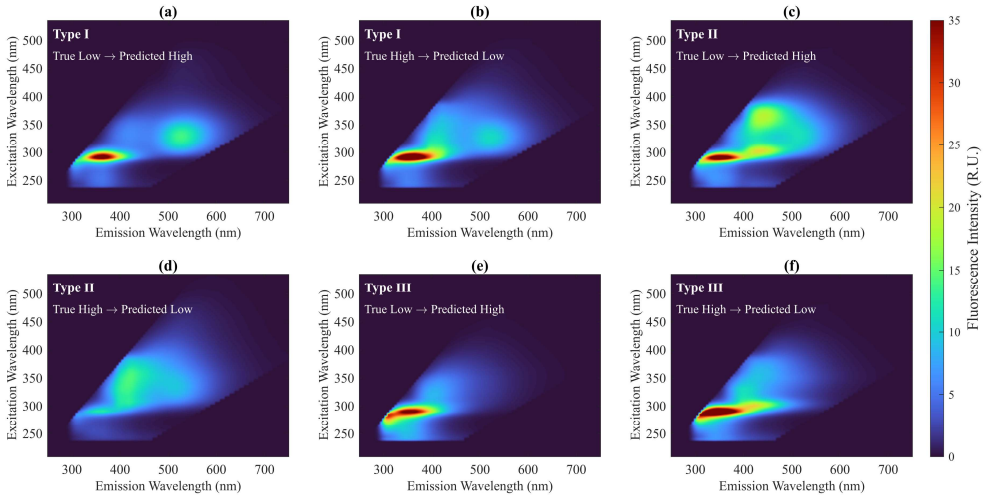


Fig. 5 Spectral characteristics of misclassified sake samples in PLS-DA model: (a–b) Type I; (c–d) Type II; (e–f) Type III, showing false positive and false negative patterns.

Type III misclassifications were dominated by amino-acid fluorescence. False positives (n = 1) exhibited weak Peak A with negligible Peaks B–C (Fig. 5e), whereas false negatives (n = 3) showed variable Peak A intensities. These findings indicate that Peak A intensity alone is insufficient for reliable discrimination. Overall, the errors originate from biochemical heterogeneity rather than model inadequacy, Future studies should consider component-wise interpretation of fluorescence features examining amino acid, Maillard, riboflavin, and humic-like emissions separately to enhance mechanistic understanding and improve model interpretability.

IV. Conclusion

This study successfully demonstrated the application of Excitation–Emission Matrix (EEM) fluorescence spectroscopy combined with chemometric modeling as a rapid, non-destructive method for classifying sake acidity. Analysis of 100 commercial sake samples revealed that Partial Least Squares Discriminant Analysis (PLS-DA) substantially outperformed Principal Component Analysis–Support Vector Machine (PCA-SVM) approaches, achieving 81% classification accuracy with balanced performance across both acidity classes (F1-scores >80%). Variable Importance in Projection (VIP) analysis identified aromatic amino acid fluorescence (Peak A) and Maillard reaction products (Peak B) as the primary discriminators, with riboflavin regions (Peak C) contributing secondarily. High-acidity sake exhibited enhanced Maillard-derived fluorescence, whereas low-acidity sake retained stronger aromatic amino acid signals. Misclassification analysis revealed that errors primarily stemmed from biochemical heterogeneity in rare spectral patterns, particularly Type II spectra (37.5% error rate), rather than fundamental model limitations. The limited representation of Type II samples (8% of dataset) highlights the importance of expanding sample diversity in future studies.

The fluorescence-based methodology offers distinct advantages over conventional titrimetric analysis, including elimination of chemical reagents, sample preservation, and potential for real-time quality monitoring. Future research should focus on implementing advanced spectral decomposition techniques to enhance classification robustness. This work establishes a foundation for broader application of fluorescence spectroscopy in fermented beverage quality evaluation.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Pornpimol Moolkaew conducted the research, performed data curation and formal analysis, developed the methodology, and prepared the original draft. Yoshito Saito supervised the research, contributed to conceptualization, project administration, and validation, and revised the manuscript. Takumi Murai assisted in data analysis and provided constructive suggestions throughout the study. Yu Obata provided consultation and contributed to data interpretation and analysis. Hideo Hasegawa supervised the overall project and contributed to project administration. All authors have read and approved the final version of the manuscript.

Funding

This work was supported in part by JSPS KAKENHI Grant Numbers JP25K22399, Japan.

References

[1] T. Asano and N. Kurose, "Mechanism of Organic Acids Production by Sake Yeast," *J. Brew. Soc. Japan.*, vol. 95, no. 4, pp. 227–234, 2000.

[2] R. Sen, K. Hope, B. Bell, and S. Lafontaine, "Evaluating the Chemical Composition and Sensory Attributes of Japanese and U.S. Sake," *ACS Food Sci. Technol.*, vol. 5, no. 7, pp. 2628–2638, July 2025.

[3] M. Ye, T. Yue, Y. Yuan, and Z. Li, "Application of FT-NIR Spectroscopy to Apple Wine for Rapid Simultaneous Determination of Soluble Solids Content, pH, Total Acidity, and Total Ester Content," *Food Bioprocess Technol.*, vol. 7, no. 10, pp. 3055–3062, Oct. 2014.

[4] U. Regmi, M. Palma, and C. G. Barroso, "Direct determination of organic acids in wine and wine-derived products by Fourier transform infrared (FT-IR) spectroscopy and chemometric techniques," *Anal Chim Acta*., vol. 732, pp. 137–144, June 2012.

[5] R.-C. Suciu, L. Zarbo, F. Guyon, and D. A. Magdas, "Application of fluorescence spectroscopy using classical right angle technique in white wines classification," *Sci Rep.*, vol. 9, no. 1, p. 18250, Dec. 2019.

[6] J. Sádecká and M. Jakubíková, "Varietal classification of white wines by fluorescence spectroscopy," *J Food Sci Technol*., vol. 57, no. 7, pp. 2545–2553, July 2020.

[7] Y. Obata, T. Murai, Y. Iida, Y. Moriguchi, and Y. Saito, "Discrimination between male-sterility and male- fertility in Japanese

cedar (Cryptomeria japonica) using fluorescence spectroscopy," *Intell., Inf. Infrastruct.*, vol. 6, no. 2, pp. 13–22, 2025.
[8] S. Ropciuc, F. Dranca, D. Pauliuc, and M. Oroian, "Honey authentication and adulteration detection using emission–excitation spectra combined with chemometrics," *Spectrochim. Acta A: Mol. Biomol. Spectrosc.*, vol. 293, 122459, May 2023.
[9] S. Boerzhijin, A. Isogai, and N. Mukai, "Impact of storage conditions on the volatile aroma compounds of aged sake," *J. Food Compos. Anal*., vol. 121, 105351, Aug. 2023.
[10] A. J. Lawaetz and C. A. Stedmon, "Fluorescence intensity calibration using the raman scatter peak of water," *Appl Spectrosc.*, vol. 63, no. 8, pp. 936–940, Aug. 2009.
[11] D. Lorente, N. Aleixos, J. Gómez-Sanchis, S. Cubero, O. L. García-Navarrete, and J. Blasco, "Recent advances and applications of hyperspectral imaging for fruit and vegetable quality assessment," *Food Bioprocess Technol.*, vol. 5, no. 4, pp. 1121–1142, May 2012.
[12] H. Abdi and L. J. Williams, "Principal component analysis, in Wiley interdisciplinary reviews: Computational statistics, vol. 2, no. 4. Chichester, U.K.: Wiley, 2010, pp. 433–459.
[13] Y. Saito, K. Itakura, M. Kuramoto, T. Kaho, N. Ohtake, H. Hasegawa, T. Suzuki, and N. Kondo, "Prediction of protein and oil contents in soybeans using fluorescence excitation emission matrix," *Food Chem.*, vol. 365, 130403, Dec. 2021.
[14] H. Chen, Y. Zhu, Y. Xie, W. Long, W. Lan, Y. She, and H. Fu, "Rapid identification of high-temperature *Daqu* Baijiu with the same aroma type through the excitation emission matrix fluorescence of maillard reaction products," *Food Control*, vol. 153, 109938, Nov. 2023.
[15] H. Yang, X. Xiao, X. S. Zhao, L. Hu, X. F. Xue, and J. S. Ye, "Study on fluorescence spectra of thiamine and riboflavin," *MATEC Web Conf.*, vol. 63, 03013, 2016.
[16] X. Xu, J. Kang, J. Shen, S. Zhao, B. Wang, X. Zhang, and Z. Chen, "EEM–PARAFAC characterization of dissolved organic matter and its relationship with disinfection by-products formation potential in drinking water sources of northeastern China," *Sci. Total Environ.,* vol. 774, 145297, June 2021.
[17] E. C. Rath, "Sake journal (Goshu no nikki)—Japan's oldest guide to brewing," Gastronomica, vol. 21, no. 4, pp. 42–50, 2021.
[18] T. Ito, H. Takahashi, T. Shiga, T. Sato, N. Nakazawa, and K. Iwano, "Analysis of the amino acid content of sake koji and estimation of the amino acid amount formed by Aspergillus oryzae in sake koji making.," *J. Brew. Soc. Japan.,* vol. 108, no. 6, pp. 453–460, 2013.
[19] A. Nose, T. Hamasaki, and M. Hojo, "Effect of humic acid in water on the coloring of Japanese sakes," *J. Brew. Soc. Japan.,* vol. 110, no. 7, pp. 525–533, 2015.
[20] F. Cosme, L. Filipe-Ribeiro, and F. M. Nunes, "Wine stabilisation: An overview of defects and treatments," 2021, pp. 1–32.
[21] E. Sikorska, T. Górecki, I. V. Khmelinskii, M. Sikorski, and D. De Keukeleire, "Monitoring beer during storage by fluorescence spectroscopy," *Food Chem.*, vol. 96, no. 4, pp. 632–639, June 2006.
[22] Y. Chen and M. D. Barkley, "Toward understanding tryptophan fluorescence in proteins," *Biochemistry*, vol. 37, no. 28, pp. 9976–9982, July 1998.
[23] H. B. Cardoso, M. Frommhagen, P. A. Wierenga, H. Gruppen, and H. A. Schols, "Maillard induced saccharide degradation and its effects on protein glycation and aggregation," *Food Chem. Adv.,* vol. 2, 100165, Oct. 2023.
[24] F. Nabi, A. Sarfaraz, R. Kama, R. Kanwal, and H. Li, "Structure-Based function of humic acid in abiotic stress alleviation in plants: A review," *Plants*, vol. 14, no. 13, 1916, Jan. 2025.