PAPER

Check for updates

Real-time prediction of shelf-life of soymilk using a surface-enhanced Raman spectroscopy (SERS) fiber and convolutional neural networks

Bezalel Adainoo, [ID] Zili Gao and Lili He*

Predicting the shelf-life of food is important for reducing food waste and ensuring consumer safety. The shelf-life of food products is predicted using models generated from data obtained from microbial, flavor, compositional and sensory analyses. However, these methods are laborious, expensive, time-consuming, and impractical for real-time analyses. In this study, a surface-enhanced Raman spectroscopy (SERS) fiber was used together with convolutional neural networks (CNNs) to develop models and predict the remaining shelf-life of soymilk during accelerated storage at 25 °C. The fiber detected the presence of different volatile organic compounds (VOCs) and varying concentrations of dimethyl sulfide during storage. In the early days of storage (days 0–5), the presence of VOCs responsible for the beany and grassy odor typical of soymilk was detected. On day 9, the presence of ketones, esters and some aldehydes was detected in the headspace. Using CNN models, the SERS spectra showed strong correlations with key quality and safety indicators including optical density ($R = 0.85$, RMSE = 0.04), pH ($R = 0.87$, RMSE = 0.32), microbial count ($R = 0.91$, RMSE = 0.69 $\log_{10}$ CFU mL^{-1}), electrical conductivity ($R = 0.92$, RMSE = 0.07 mV), particle size ($R = 0.94$, RMSE = 212.59 nm), and zeta-potential ($R = 0.94$, RMSE = 1.28 mS cm^{-1}). The SERS spectra also showed strong correlations with the remaining shelf-life ($R = 0.95$, RMSE = 1.30 days). Separate spectra were used to externally validate the remaining shelf-life and microbial count models. The results demonstrated strong predictive performance, with the model achieving accurate predictions for the remaining shelf-life and microbial count. These findings support the potential of the SERS fiber–CNN approach for practical shelf-life prediction. More tests are needed for different food products and conditions.

Sustainability spotlight

The developed surface-enhanced Raman spectroscopy (SERS) fiber combined with convolutional neural network models for real-time shelf-life prediction promotes sustainability in the food system by introducing a rapid, non-invasive and data-driven approach for predicting soymilk shelf-life which can be extended to other food products. This study offers a solution that enables real-time and accurate monitoring of food quality and safety indicators. The technology can potentially allow food producers and retailers to better assess the actual freshness and safety of food products, reducing unnecessary premature disposal, hence minimizing food waste. By maximizing the usable life of food products and ensuring they are only discarded when truly unsafe or of unacceptable quality, this approach not only helps conserve resources and reduce environmental burden but also supports food security and economic efficiency throughout the supply chain.

1   Introduction

Accurately predicting the shelf-life of food is crucial for reducing food waste and boosting consumer confidence in the food industry. Recent studies have established an inverse relationship between shelf-life and food waste primarily because consumers discard packaged food even before it is spoiled.[1,2] This significantly contributes to food waste. Conventionally, food analysis methods such as microbial analysis, sensory analysis, flavor profile analysis using gas chromatography-mass spectrometry (GC-MS), and compositional analysis are used to obtain data during storage of food products, which are then used to develop various prediction models to determine the shelf-life of these food products.[3] However, while these methods are accurate, they are laborious, expensive, and time-consuming, making them impractical for real-time monitoring of food shelf-life. Monitoring quality and safety changes in food products in real-time is essential for safeguarding the food system and quick decision-making to protect consumer health. Several studies have demonstrated that headspace volatile organic compounds (VOCs) provide real-time insights into food

Department of Food Science, University of Massachusetts, Amherst, MA 01003, USA. E-mail: lilihe@umass.edu

spoilage.[4] Other studies have explored the combination of GC-MS headspace analysis and neural networks for predicting the shelf-life of food products.[5,6]

Surface-enhanced Raman spectroscopy (SERS) is an analytical technique that involves the use of a roughened surface with a nanoparticle coating that enhances the Raman signal of analytes when excited with a laser. SERS has effectively been utilized to investigate VOCs in various samples including distinguishing healthy and pest-infested plants based on their VOCs, identification of disease-related VOC biomarkers and detection of metabolic products in the headspace bacterial cultures.[7-9] SERS is useful for these analyses because it is simple, inexpensive, and rapid. In previous studies, a SERS-fiber has been applied to study the changes in the headspace VOCs of whole milk and arugula leaves during storage.[10,11] The SERS fiber is a gold-coated stainless steel wire on which molecules of the VOCs adsorb when inserted into the headspace of a sample. Spectral data can then be acquired from the SERS fiber which provides insight into the VOCs that are present in the sample headspace.

Traditionally, chemometric data analyses such as principal component analysis (PCA) and partial least squares (PLS) are used to process SERS spectral data. Chemometric analyses make it possible to preprocess spectral data to enhance data quality and extract as much valuable information as possible.[12] However, these data analysis methods can cause overfitting which increases the risk of false discoveries and reduces generalizability of models to new data.[13] Some studies have explored the combination of spectral data and chemometric analyses to predict the shelf life or quality of different foods.[11,14] These methods may also struggle to capture the nonlinear relationships and high-dimensional features commonly present in SERS data, limiting their effectiveness for real-time applications such as food spoilage prediction. On the other hand, convolutional neural networks (CNNs) perform very well at handling raw, high-dimensional spectral data, can extract valuable features directly from raw spectra, and provide models that are typically easier to interpret.[15,16] As a result, CNNs are increasingly being used for spectral data analysis in both food quality monitoring and biomedical applications.[17]

In our most recent work, the combination of spectra from the SERS fiber and CNN model showed great promise in detecting the changes in the headspace of whole milk during storage and correlations with key quality and safety indicators.[10] However, since whole milk spoils quickly, we could not explore the potential of the SERS fiber to predict the remaining shelf-life. The aim of this current research was to predict the remaining shelf-life of soymilk in accelerated storage (25 °C) using headspace spectra from the SERS fiber and CNN models. Soymilk was used for this study because its spoilage is a gradual process,[18] which would allow for collection of sufficient headspace data over a two-week period to allow for the prediction of the remaining shelf-life during accelerated storage.

2 Materials and methods

2.1 Materials

Polished 304 stainless steel wire (ø 0.3 mm) was purchased from Vigan *via* Amazon. Pure ethanol (200 Proof) was purchased from

Decon Labs Inc. (King of Prussia, PA, USA). Hydrochloric acid (36.5 to 38.0% w/w) was purchased from Fisher Scientific (Fair Lawn, NJ, USA). Hydrogen tetrachloroaurate(III) hydrate was purchased from Sigma-Aldrich (St. Louis, MO, USA), and 2% stock solution was prepared using distilled water and further diluted to a 0.1% solution which was used in the SERS fiber fabrication. Plain, unsweetened soymilk (Brand A) from different manufacturing batches with at least one month left on its best by date was purchased from two different local shops at different times: Stop & Shop and Walmart (Hadley, MA). The soymilk from Stop & Shop was purchased in March 2024 (Brand A1) to obtain the training data and the soymilk from Walmart was purchased in February 2025 (Brand A2) to obtain the test data for the models developed in this study. In addition, two other brands of plain, unsweetened soymilk (Brands B and C) were purchased in October 2025 to validate the performance of the shelf life prediction model.

2.2 SERS fiber fabrication

The SERS-fiber used in this experiment was fabricated using the method described in our previous work[10] with some modifications. The stainless steel wire was cut into 8 cm pieces and washed with distilled water followed by washing with absolute ethanol using an ultrasonic bath (Branson CPX 2800H Digital Heated Ultrasonic Cleaner, Branson Ultrasonics Corporation, Danbury, CT, USA) at frequency of 40 kHz and a power output of 110 W for 5 minutes at 25 °C. This was followed by air drying for about 10 minutes then the dried wire pieces were etched in concentrated hydrochloric acid for 30 minutes to enhance the gold coating. The hydrochloric acid was poured off the wire pieces and the wire pieces were washed with distilled water followed by washing with absolute ethanol for 5 minutes using the ultrasonic bath. The washed etched wire pieces were then air dried by blowing air. The dried etched wire pieces were immersed in 0.1% (v/v) hydrogen tetrachloroaurate(III) solution at room temperature for 45 minutes to allow for the iron–gold replacement reaction to take place, which resulted in the gold coating on the wire. The gold-coated SERS fiber was stored in a glass vial for use.

2.3 Experimental setup for headspace analyses using the SERS fiber

Soymilk samples were poured to half-fill 15 ml glass vials with PTFE/silicone septa and open-top polypropylene caps. The capped vials were stored at room temperature (25 ± 2 °C) over a two-week period. Two vials were taken and analyzed for headspace volatiles and other key quality indicators at each time point (days 0, 3, 5, 7, 9, 11 and 13). Two SERS fibers were inserted into the headspace of each soymilk glass vial with the aid of syringe needles and incubated at room temperature for 30 minutes. The SERS fibers were then immobilized onto a plain glass slide and SERS spectra were acquired using a Raman microscope. This experimental setup was used for both sets of soymilk samples used for training and testing the models.

2.4 Raman instrumentation

A DXR Raman microscope system (Thermo Fisher Scientific, Madison, WI, USA) equipped with a 780 nm laser source

(maximum at 24 mW) was used to collect the SERS spectra in this study. The Raman instrument was controlled with OMNIC™ software (version 9.1). The spectra were collected using a 20× objective lens and a detection range from 400 to 3200 cm^{-1}. The measurement was carried out over a 2-second exposure time and a 3 mW laser power. Sixty spectra were collected for each SERS fiber. For the samples used to develop the models, 20 spectra were selected for each fiber, making a total of 80 spectra for each time point for data processing and analyses. For the samples used to validate the models, fifty spectra were selected for each fiber (making a total of 200 spectra for each time point) for data processing and analyses.

2.5 Data preprocessing

Prior to analysis, all spectra were preprocessed using z-score normalization. For each spectrum, the mean and standard deviation across all Raman shift positions were calculated, and normalization was performed by subtracting the mean and dividing by the standard deviation. This standardized each spectrum to have a mean of 0 and a standard deviation of 1, improving comparability across samples. To further reduce the baseline drift and high-frequency noise, a Norris second derivative filter[19] was applied to the normalized spectra. The resulting preprocessed data were used to train the CNN models.

2.6 CNN model establishment and evaluation

A one-dimensional convolutional neural network (1D-CNN) model was developed to analyze headspace SERS spectra for both classification and regression tasks, following a similar approach described in our previous work.[10] The architecture consisted of two convolutional layers, each followed by batch normalization and max pooling to reduce dimensionality while retaining key spectral features. These were followed by fully connected layers with ReLU activations to extract higher-level representations. The output layer was task-specific: producing class probabilities for classification and a single continuous output for regression.

Data were split into training and testing sets, converted to tensors, and processed using PyTorch. Cross-entropy loss and mean squared error (MSE) were used for classification and regression tasks, respectively, with optimization performed using the Adam algorithm. Full-batch training was applied, and model performance was evaluated using accuracy for classification, and R and RMSE for regression. A confusion matrix and t-SNE were also used to visualize the classification results.

The developed CNN model was evaluated using a new set of headspace SERS spectra from a separate soymilk sample during its initial storage period. The prediction accuracy of the remaining shelf-life and microbial counts of this new sample was determined by comparing to the actual days and microbial counts.

2.7 Analyses of key quality indicators

2.7.1 Microbiological analysis. Microbiological analysis of the soymilk samples was conducted at each time point (days 0, 3, 5, 7, 9, 11 and 13) to obtain the total plate count. Serial dilutions of the soymilk samples were performed using peptone water and 200 μl of the dilution was plated on tryptic soy agar by spread plating and incubated at 37 °C for 24 to 48 hours. The total plate count (CFU ml^{-1}) was calculated from the colony counts, and $\log_{10}$ conversions of the total plate counts were calculated and used for further analyses.

2.7.2 pH analysis. The pH of the soymilk samples was monitored at each time point during the storage period using a digital pH meter (Fisher Scientific accumet AE150) at room temperature (25 °C). The measurements were performed in triplicate.

2.7.3 Optical density analysis. To determine the optical density (OD$_{600}$) of the soymilk samples, 10-fold dilutions of the samples were prepared. Seven aliquots of 200 μl each were dispensed into wells of a 96-well plate, using 200 μl distilled water as the reference. Absorbances at 600 nm were measured with a UV-Vis spectrophotometer (Molecular Devices, SpectraMax M2, San Jose, CA, USA), and the readings were recorded as the optical density.

2.7.4 Particle size, zeta potential and electrical conductivity analysis. The particle size of the soymilk samples at each time point was measured by dynamic light scattering (Zetasizer Nano ZS, Malvern Instruments Inc., Malvern, UK) as described by Li & McClements[20] using the same dilutions described in Section 2.7.3. The refractive index of the dispersion medium (water) used for the determination of the particle size was 1.33. The analyses were performed at 25 °C with a refractive index of 1.35 and a count rate of 76.5 kcps.

Zeta potential and electrical conductivity were measured by electrophoresis using a Zetasizer Nano ZS (Malvern Instruments Inc., Malvern, UK). The dilution and instrument settings were the same as those used in the particle size analysis. For each time point, at least three measurements were taken.

2.8 Statistical analysis

The averages and standard deviations of the results from the microbiological, pH, optical density, particle size, zeta potential and electrical conductivity were calculated and plotted on graphs. Furthermore, the microbiological, pH, optical density, particle size, zeta potential and electrical conductivity data were analyzed for significant differences between the time points by analysis of variance (one-way ANOVA) and Tukey's test ($p < 0.05$). The statistical analyses were done using OriginPro 2024b version 10.1.5 software (Origin Lab Inc., Northampton, MA, USA).

3 Results and discussion

3.1 Spectral changes in the headspace during storage

The results from the study show that during the two-week storage period, there were spectral changes in the headspace of the soymilk samples indicating changes in VOCs (Fig. 1). On days 0–5, there were broad bands between 800 cm^{-1} and 1000 cm^{-1} which indicate overlapping vibrational modes from aromatic rings, C–C stretching and C–H bending.[21] The molecules with these vibrational modes which may have been
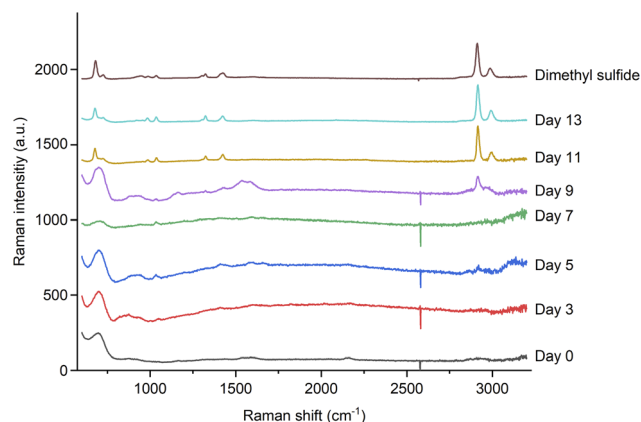
Fig. 1 Headspace SERS spectra of soymilk samples during storage at 25 °C.

Fig. 2 t-SNE visualization of the classification of the SERS spectra.

detected by the SERS fiber include hexanal, 2-pentylfuran and 1-octen-3-ol, which are also responsible for the beany, grassy odor observed on days 0–5.[22] Also, starting on day 3, a weak peak was observed around 1035 cm^{-1}, which increased gradually on day 5 and day 7. This may indicate the presence of aromatic thiols such as thiophenol and alkyl substituted thiophenols in the headspace.[23,24] Also, this peak may indicate the C–S stretching in dimethyl sulfide.[10,25] No beany, grassy odor was observed on day 7, which is evident in the spectrum as the broad band between 800 cm^{-1} and 1000 cm^{-1} present on days 0–5 was absent. On day 9, the broad bands from 1110 to 1200 cm^{-1} and 1480 to 1660 cm^{-1} are consistent with the C–O stretching and C=C stretching respectively.[21] These broad bands suggest the presence of other volatile organic compounds such as esters, some aldehydes, ketones, and carboxylic acids in the headspace of the soymilk samples by day 9. Notably, on day 9, the soymilk samples had a fruity smell which was perceivable by smell, suggesting the presence of esters and some aldehydes (like nonanal) in the headspace. After day 9, other peaks, which are consistent with the peaks in dimethyl sulfide became more pronounced indicating an increase in the concentration of dimethyl sulfide on days 11 and 13 when the samples were completely spoiled. On day 11, the peaks at 676, 726, 983, 1035, 1323, 1424, 2915, and 2990 cm^{-1} became more apparent, and these peaks increased in intensity on day 13 further showing an increased concentration of dimethyl sulfide in the headspace as the soymilk samples spoiled.[10,25]

The spectral data and observations are consistent with the formation and release of volatile organic compounds from the soymilk samples reported in other studies.[26] During soymilk spoilage, the perceived odor was consistent with the fruity and cooked bean odor indicative of the presence of aldehydes and esters (like 2-heptenal, nonanal, heptanal, ethyl acetate, and 2,4-heptadienal) and spoiled odor consistent with the presence of sulfides from the breakdown of sulfur-containing amino acids like cysteine.[27] This shows that the SERS fiber, which only costs $0.07 to make can obtain headspace data consistent with data obtained from GC-MS in a much shorter time (in 2 minutes) than the analysis time of GC-MS (1–2 hours).
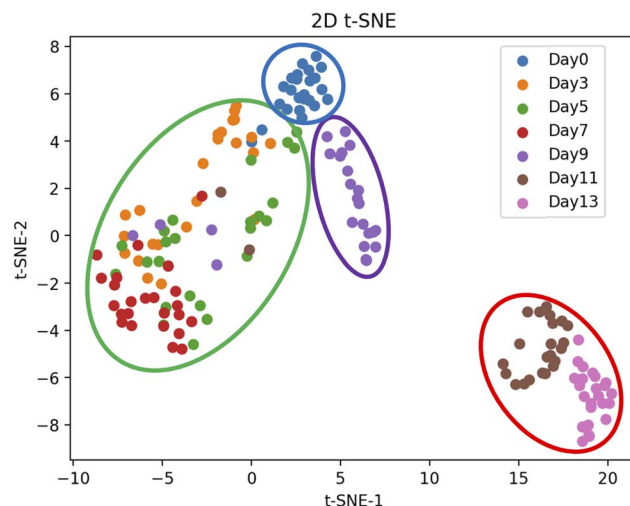
3.2 Classification of SERS spectra using the CNN model

The classification of the SERS spectra was done using CNN and non-CNN (Mahalanobis distance-based classification using principal component analysis for dimensionality reduction) models. Seventy percent of the spectral data from the first batch of the soymilk samples was used to train the model and the remaining 30% was used to test the models. The results were visualized with t-distributed Stochastic Neighbor Embedding (t-SNE) (Fig. 2) and a confusion matrix (Fig. 3) was used to visualize the accuracy of the classification on each of the test days. The CNN model achieved a classification accuracy of 84%, which was significantly higher than the classification accuracy yielded from non-CNN analysis (36%). While the non-CNN data analysis is suitable for classification, the CNN data analysis achieved better test accuracy because it could better extract
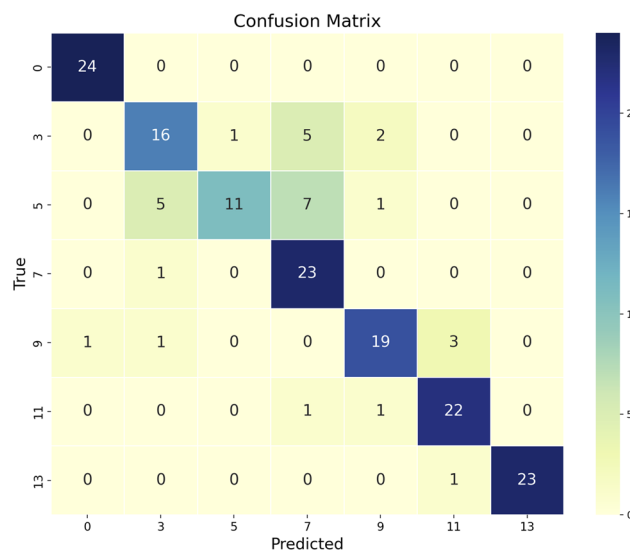

Fig. 3 Confusion matrix showing the test accuracy on each test day during the storage period.

spectral features and reduce the effect of noise to achieve better results.[28,29] Furthermore, the *t*-SNE plot shows that the CNN model could clearly distinguish between the headspace spectra of soymilk on day 0 in the blue ellipse and those on days 3 and 5 in the green ellipse even though there are minor observable differences between the spectra on those days in Fig. 1. The day 7 data points were more separated with a few overlapping with day 3 and 5 clearly. The day 9 data points were well separated and clustered, when the samples started to develop fruity and spoiled smells. The headspace spectra from days 11 and 13 were distinctly separated from the other samples, as the soymilk developed increasingly strong sulfurous odor during this period (Fig. 2).

The confusion matrix (Fig. 3) shows that ,overall, the CNN model could accurately classify most of the spectral data. All the spectral data used to test the model on day 0 were accurately classified. The model, however, misclassified some of the test data on days 3 and 5. Some of the day 3 spectra were misclassified as spectra taken on day 5, day 7 and day 9, whereas some of the day 5 spectra were misclassified as spectra taken on day 3, day 7 and day 9 (Fig. 3). These misclassifications could be attributed to the higher degree of similarity in the spectral data collected on days 3, 5, and 7, a pattern that is also apparent in the *t*-SNE plot (Fig. 2). The spectra collected on day 7, day 9, day 11 and day 13 were mostly classified accurately.

3.3 Analyses of key quality indicators

In addition to the headspace VOC analyses, other quality indicators of soymilk including pH, microbial count, optical density, particle size, electrical conductivity and zeta-potential were assessed during storage at 25 °C (Fig. 4). These quality indicators were evaluated to further understand how the changes in the headspace VOCs as detected by the SERS-active fiber relate to these known quality indicators.

The pH of the soymilk product tested in this study was 8.16 ± 0.02 on day 0 (Fig. 4A), which is very alkaline. During soymilk production, an alkaline pH is known to be useful in the extraction of soymilk from soybeans as it yields a higher
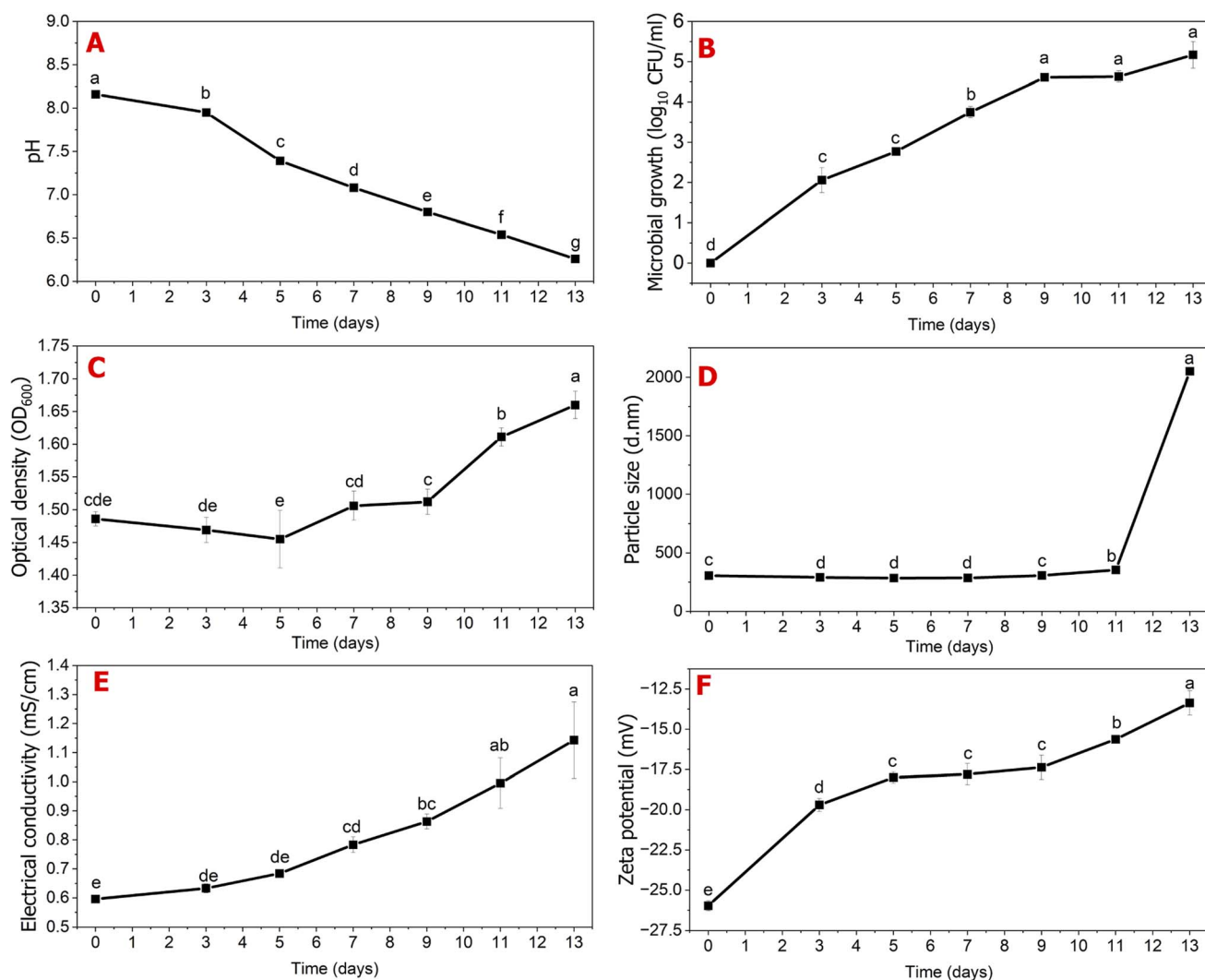

Fig. 4 Physicochemical properties and microbial count: (A) pH; (B) microbial growth; (C) optical density; (D) particle size; (E) electrical conductivity and (F) zeta potential of soymilk samples during storage at 25 °C.

extraction of protein, fat, and solids in soymilk and reduces the formation of beany flavors in soymilk.[30,31] Since the pH was more alkaline on day 0, it did not support any microbial growth (Fig. 4B). Furthermore, soymilk is known to contain antimicrobial compounds like isoflavones which could have also inhibited microbial growth.[32] Nonetheless, the microbial count increased almost linearly over time as the pH decreased. During storage, soymilk became increasingly acidic due to enzymatic and microbial activities, particularly microbial fermentation, which led to various chemical changes and the production of acids. Hence, the pH of the soymilk samples decreased and reached 6.26 ± 0.01 on day 13 (Fig. 4A).

Optical properties such as turbidity and particle size are known to be good indicators of soymilk quality.[33,34] Turbidity is one of the important optical properties that can be used to determine soymilk quality, stability, and protein content.[31] In this study, turbidity was measured during soymilk storage as optical density at 600 nm (OD_{600}). The results show that the OD_{600} increased during the two-week storage from 1.468 ± 0.011 on day 0 to 1.660 ± 0.021 on day 13 (Fig. 4C). This pattern of change is consistent with changes in optical density of emulsions and suspensions as they become unstable.[35] These findings show that during the storage period, there were changes in the soymilk quality and the samples became unstable over time. The increase in optical density was also consistent with the visual observations made of the viscosity of the soymilk samples. The samples became more viscous during the storage period. Generally, smaller, uniform particle sizes indicate better soymilk quality and stability. According to Fan et al.,[36] the particle size of soymilk increases steadily from about 600 nm to about 900 nm over a 100-day period at 25 °C. However, the results obtained in this study show that the particle size gradually reduced from 309.5 ± 5.4 nm on day 0 to a lowest of 284.8 ± 0.9 nm on day 5 after which the particle size increased to 354.5 ± 1.5 nm on day 11 and then increased sharply to 2051.3 ± 10.1 nm on day 13 (Fig. 4D). The reduction in the particle size between day 0 and day 5 may have resulted from the microbial breakdown of the soymilk particles, while the increase in the particle size after day 5 may have been due to the agglomeration of the particles caused by the changes in pH and ionic strength.[37,38]

Electrical conductivity is a measure of the ability of a material to conduct electric current. During food spoilage, as the food particles break down, they release more free ions into solution, thereby increasing the electrical conductivity. Due to this strong relationship between electrical conductivity and food spoilage, it has been successfully used to evaluate the quality of various foods including soymilk, tofu, and pomegranate juice, among others.[35,39,40] During the storage period, the electrical conductivity increased from 0.596 ± 0.001 mS cm^{-1} on day 0 to 1.143 ± 0.131 mS cm^{-1} on day 13 as the spoilage progressed (Fig. 4E). This clearly shows that over time, more free ions were released from the breakdown of the soymilk particles into solution, indicating spoilage during storage.

Furthermore, zeta-potential was used as a measure of the stability of the soymilk during storage. During storage, the soymilk became less stable, which is evident in the zeta-

potential measured (Fig. 4F). The samples had a zeta-potential of -25.97 ± 0.31 mV on day 0, which is very close to zeta-potential values consistent with stable colloids.[41] As the soymilk spoiled, the zeta-potential increased to -13.37 ± 0.75 mV on day 13, when the soymilk samples had become unstable and separated into two phases. These changes occurred as a result of the changes in the pH and surface charge, which resulted in agglomeration and consequent instability.[42]

3.4 Correlations between headspace spectra from the SERS fiber and key quality and safety indicators

Regression analyses between the headspace spectra obtained with the SERS fiber and the key quality and safety indicators measured in this study showed strong correlations between the headspace spectra and all the quality indicators (Table 1). The changes in the spectra during storage had a correlation coefficient (R) of 0.85 (RMSE = 0.04) with the changes in the optical density. This is because, as the soymilk spoiled and released VOCs into the headspace, the samples showed an increase in the optical density at a similar rate. Similarly, the change in pH showed strong correlations with the SERS spectra (R = 0.87, RMSE = 0.32). Since a decrease in pH is a good indicator of microbial spoilage in soymilk,[43] its strong correlation with the headspace SERS spectra obtained with the SERS fiber demonstrates that this method is reliable for monitoring spoilage. Furthermore, the microbial count even showed a stronger correlation with the headspace SERS spectra (R = 0.91, RMSE = 0.69 log$_{10}$ CFU ml^{-1}). Generally, in predictive microbiology and food safety, an error margin of ± 1 log is considered acceptable for model performance and to ensure the protection of public health.[44] Since the RMSE of the microbial count regression analysis is within the acceptable error margin, it demonstrates the reliability of combining the SERS fiber and CNN modeling

Table 1 Correlation coefficient (R), and root mean square error (RMSE) from the regression analyses between the headspace spectra and measured key quality and safety indicators using convolutional neural networks (CNNs)

Quality indicator	R	RMSE
Optical density	0.85	0.04
pH	0.87	0.32
Microbial count (log$_{10}$ CFU ml^{-1})	0.91	0.69
Electrical conductivity (mS cm^{-1})	0.92	0.07
Particle size (nm)	0.94	212.59
Zeta potential (mV)	0.94	1.28
Remaining shelf life (days)	0.95	1.30

Table 2 Prediction of soymilk Brand A2 remaining shelf life using the CNN model

Day	True shelf life (days)	Predicted shelf life (days)	Predicted standard deviation (days)
0	9	8.56	1.33
3	6	6.16	1.23
5	4	5.07	1.10

Table 3 Prediction of soymilk Brand A2 microbial count using the CNN model

Day	True microbial count ($\log_{10}$ CFU ml^{-1})	Predicted microbial count ($\log_{10}$ CFU ml^{-1})	Predicted standard deviation ($\log_{10}$ CFU ml^{-1})
0	0.00	0.12	0.24
3	2.55	2.14	0.58
5	2.82	2.38	0.55

Table 4 Validation of shelf life model performance using two other soymilk brands

Day	True shelf life (days)	Predicted shelf life (days)	Predicted standard deviation (days)
Brand B			
0	9	7.28	1.38
3	6	6.35	0.89
5	4	5.14	1.18
Brand C			
0	9	7.52	0.93
3	6	6.41	1.05
5	4	4.97	1.22
Brand A1 + A2 + B (train), Brand C (test)			
0	9	8.23	1.09
3	6	6.47	0.9
5	4	4.66	1.08
Brand A1 + A2 + C (train), Brand B (test)			
0	9	8.31	1.16
3	6	6.41	0.93
5	4	4.38	0.77

for monitoring and predicting soymilk safety in real-time. The headspace SERS spectra also had strong correlations with electrical conductivity ($R = 0.92$, RMSE = 0.07 mV), which is an important food spoilage indicator. The particle size showed strong correlations ($R = 0.94$, RMSE = 212.59 nm) (Table 1) with the headspace SERS spectra as well; however, the large RMSE is a result of the significant increase from 354.5 ± 1.5 nm on day 11 to 2051.3 ± 10.1 nm on day 13 (Fig. 4D). Zeta-potential had a similarly strong correlation ($R = 0.94$, RMSE = 1.28 mS cm^{-1}) with the headspace spectra during storage.

In order to use the headspace SERS spectra to predict the remaining shelf-life of soymilk, a regression analysis was performed with day 9 as the expiry date since that is the day on which the microbial load reached the acceptable limit of 10^4 CFU ml^{-1}.[45] Hence, the remaining shelf-life on the respective days was as follows: day 0 = 9 days, day 3 = 6 days, day 5 = 4 days, day 7 = 2 days, day 9 = 0 days, day 11 = −2 days, day 13 = −4 days. Days 11 and 13 had a negative remaining shelf-life to indicate how many days had passed beyond the shelf-life. Regression analysis between the headspace SERS spectra acquired from the SERS fiber and remaining shelf-life shows that there is a strong correlation ($R = 0.95$, RMSE = 1.3 days) between the two factors. This demonstrates the great potential of the SERS fiber and CNN modeling to predict the remaining shelf-life reliably and accurately in real-time.

3.5 Evaluation of the prediction accuracy of the remaining shelf-life and microbial count of a separate sample

To evaluate the CNN model's ability to predict the freshness of new samples reliably and accurately in real-time, a separate batch of soymilk (Brand A2) was purchased and tested during the early storage days. Headspace SERS spectra were collected from these new soymilk samples during the initial storage period, and the previously developed model was used to predict the remaining shelf-life at specific intervals (9, 6, and 4 days remaining). Table 2 compares the actual remaining shelf-life with the predicted values derived from these spectra. The model accurately predicted the remaining shelf-life, retaining a consistent standard deviation (1–1.5 days) across predictions (Table 2). This demonstrates that the CNN model can effectively utilize spectral data obtained during early and fresh storage days to determine the remaining shelf-life, confirming the practical applicability of the SERS fiber and model in real-world scenarios. Additionally, the CNN model's capability to predict the microbial count was also evaluated using the same set of headspace SERS spectra (Table 3). The microbial count during storage was effectively predicted with standard deviations in the range of 0.24–0.55 $\log_{10}$ CFU ml^{-1} using only spectral data collected at the respective storage intervals.

Two other brands (Brands B and C) were later obtained, and the headspace SERS spectra were collected under the same conditions as those used for soymilk Brand A. The headspace SERS spectra from Brands B and C were used to validate the performance of the shelf life prediction model. The results show that the model predicted the Brand B samples to have 7.28 ± 1.38 days and the Brand C samples have 7.52 ± 0.93 days left on their shelf lives (Table 4).Overall, the model performed well in predicting the remaining shelf life of Brand B with an error margin of 0.89 to 1.38 days and that of Brand C with an error margin of 0.93 to 1.22 days (Table 4). These error margins are very close to those obtained when Brand A2 was used to test the model built using Brand A1 (Table 1). To further confirm the generalizability of the model, data from brands A1, A2 and B were put together to train the model and tested with Brand C data. Similarly, data from brands A1, A2 and C were put together to train the model and tested with Brand B data. The results clearly show that the performance of the model was consistent with the results obtained from only Brand A data (Table 2). Furthermore, although these soymilk brands had slightly different protein contents (Brand A = 8 g per 240 ml serving, Brand B = 7 g per 240 ml serving, Brand C = 9 g per 240 ml serving), the changes in their headspace VOCs during

accelerated storage were similar and yielded similar shelf life prediction results.

4 Conclusion

This study aimed at testing the application of the SERS-fiber and CNN modeling for predicting the shelf-life of soymilk in real-time. The findings from this study show that, using the headspace spectra collected from the SERS-fiber method described in this study together with CNN modeling, the changes in the quality and safety of packaged soymilk can be monitored in real-time. The SERS fiber detected the presence of VOCs that give soymilk its characteristic beany, grassy odor on days 0–5. The VOC profile changed to indicate the presence of ketones, esters, and some aldehydes on day 9 which gave the soymilk samples a fruity odor consistent with the physical observations made. Also, dimethyl sulfide was the predominant VOC detected on days 11 and 13, by which time the soymilk was completely spoiled. Also, the SERS fiber detected the increase in the intensity of dimethyl sulfide during the storage period. The changes in the headspace spectra showed strong correlations with optical density, pH, microbial count, electrical conductivity, particle size, zeta potential and remaining shelf-life. Furthermore, the CNN model performed very well in predicting the remaining shelf-life and microbial counts of separate sets of soymilk samples, emphasizing the capability of the combination of the SERS fiber and CNN modeling to predict the shelf-life of soymilk. Further studies are needed to further explore this technology in real-time monitoring of shelf-life of other food products.

Conflicts of interest

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Data availability

The data will be made available on request.

Acknowledgements

References

1 A. Spada, A. Conte and M. A. Del Nobile, *J. Clean. Prod.*, 2018, **172**, 3410–3414.

2 R. Neff, E. Broad Leib, A. Khan and D. Gunders, *Consumer Perceptions of Food Date Labels: 2025 National Survey*, 2025.

3 F. Cui, S. Zheng, D. Wang, X. Tan, Q. Li, J. Li and T. Li, *Compr. Rev. Food Sci. Food Saf.*, 2023, **22**, 1257–1284.

4 H. Lin, H. Jiang, S. Y. S. S. Adade, W. Kang, Z. Xue, M. Zareef and Q. Chen, *Crit. Rev. Food Sci. Nutr.*, 2023, **63**, 8226–8248.

5 B. Vallejo-Cordoba, G. E. Arteaga and S. Nakai, *J. Food Sci.*, 1995, **60**, 885–888.

6 Y. Zhang, D. Zhu, X. Ren, Y. Shen, X. Cao, H. Liu and J. Li, *Food Chem.*, 2022, **394**, 133526.

7 J. Kelly, R. Patrick, S. Patrick and S. E. J. Bell, *Angew. Chem., Int. Ed.*, 2018, **57**, 15686–15690.

8 A. Fernández-Lodeiro, M. Constantinou, C. Panteli, A. Agapiou and C. Andreou, *ACS Sens.*, 2025, **10**, 621.

9 J. Park, J. A. Thomasson, C. C. Gale, G. A. Sword, K. M. Lee, T. J. Herrman and C. P. C. Suh, *ACS Omega*, 2020, **5**, 2779–2790.

10 B. Adainoo, Z. Gao and L. He, *ACS Food Sci. Technol.*, 2025, **5**, 1504–1512.

11 X. Du, H. Chen, Z. Zhang, Y. Qu and L. He, *Postharvest Biol. Technol.*, 2021, **175**, 111410.

12 A. González-Casado, A. María, J. Carvelo, R. González-Domínguez, A. Sayago and Á. Fernández-Recamales, *Foods*, 2022, **11**, 3940.

13 H. Martens, *J. Chemom.*, 2015, **29**, 563–581.

14 M. Xiao, Y. Chen, F. Zheng, Q. An, M. Xiao, H. Wang, L. Li and Q. Dai, *npj Sci. Food*, 2023, **7**, 1–8.

15 N. M. Peleato, *Sci. Rep.*, 2022, **12**, 1–12.

16 R. Gariso, J. P. L. Coutinho, T. J. Rato and M. S. Reis, *Anal. Chim. Acta*, 2025, **1347**, 343766.

17 Y. Dong, J. Hu, J. Jin, H. Zhou, S. Jin and D. Yang, *TrAC, Trends Anal. Chem.*, 2024, **180**, 117974.

18 L. Fan, Y. Duan, Z. Huang, D. Zhao, L. Zhao, W. He, X. Zhang, M. Li, Y. Lin and Y. Chen, *Food Sci. Nutr.*, 2024, **12**, 1973–1982.

19 Y. Yang, T. Pan, J. Zhang, Y. Yang, T. Pan and J. Zhang, *Am. J. Anal. Chem.*, 2019, **10**, 143–152.

20 S. Li and D. J. McClements, *Food Hydrocoll.*, 2023, **145**, 109126.

21 H. G. M. Edwards, in *Handbook of Vibrational Spectroscopy*, John Wiley & Sons, Ltd, 2006, pp. 1838–1871.

22 X. Feng, Y. Zhu and Y. Hua, *Food Chem.: X*, 2023, **20**, 100892.

23 L. Chen, D. L. Capone and D. W. Jeffery, *Molecules*, 2019, **24**, 2472.

24 F. Madzharova, Z. Heiner and J. Kneipp, *J. Phys. Chem. C*, 2020, **124**, 6233–6241.

25 J. K. Lim, I. H. Kim, K. H. Kim, K. S. Shin, W. Kang, J. Choo and S. W. Joo, *Chem. Phys.*, 2006, **330**, 245–252.

26 W. Zhang, X. Liu, Z. Yang, H. Song, Y. Zhang and Y. Jin, *J. Food Sci. Technol.*, 2018, **55**, 1591–1598.

27 Z. Hao, X. Zhang, X. Peng, X. Shi, R. Wang and S. Guo, *Food Res. Int.*, 2023, **164**, 112407.

28 J. Walsh, A. Neupane and M. Li, *Spectrochim. Acta, Part A*, 2024, **311**, 124003.

29 P. Ravichandran, S. Viswanathan, S. Ravichandran, Y. J. Pan and Y. K. Chang, *Cereal Chem.*, 2022, **99**, 907–919.

30 X. Li, X. Liu, Y. Hua, Y. Chen, X. Kong and C. Zhang, *RSC Adv.*, 2019, **9**, 2906–2918.

31 S. K. Giri and S. Mangaraj, *Food Eng. Rev.*, 2012, **4**, 149–164.

32 R. P. A. Dhayakaran, S. Neethirajan, J. Xue and J. Shi, *LWT–Food Sci. Technol.*, 2015, **63**, 859–865.

33 D. A. Murugkar, *J. Food Sci. Technol.*, 2015, **52**, 2886–2893.

34 A. M. Nik, S. Tosh, V. Poysa, L. Woodrow and M. Corredig, *Food Res. Int.*, 2008, **41**, 286–294.

35 Z. B. Wang, P. Shan, D. Q. Wei, S. J. Hao, Z. Zhang, S. X. Li and J. Xu, *J. Pharm. Sci.*, 2021, **110**, 2416–2422.

36 L. Fan, Y. Duan, Z. Huang, D. Zhao, L. Zhao, W. He, X. Zhang, M. Li, Y. Lin and Y. Chen, *Food Sci. Nutr.*, 2023, **12**, 1973.

37 A. M. Petenate and C. E. Glatz, *Biotechnol. Bioeng.*, 1983, **25**, 3049–3058.

38 L. Sivanandan, R. T. Toledo and R. K. Singh, *Int. J. Food Prop.*, 2010, **13**, 580–598.

39 S. S. Yu, H. S. Ahn and S. H. Park, *Sens. Actuators, A*, 2023, **352**, 114202.

40 H. Darvishi, M. H. Khostaghaza and G. Najafi, *J. Saudi Soc. Agric. Sci.*, 2013, **12**, 101–108.

41 D. J. Pochapski, C. Carvalho Dos Santos, G. W. Leite, S. H. Pulcinelli and C. V. Santilli, *Langmuir*, 2021, **37**, 13379–13389.

42 S. Ferraris, M. Cazzola, V. Peretti, B. Stella and S. Spriano, *Front. Bioeng. Biotechnol.*, 2018, **6**, 369077.

43 X. Ma, X. Hu, L. Liu, X. Li, Z. Ma, J. Chen and X. Wei, *Food Sci. Nutr.*, 2016, **5**, 123.

44 K. Koutsoumanis, A. Alvarez-Ordóñez, D. Bolton, S. Bover-Cid, M. Chemaly, R. Davies, A. De Cesare, L. Herman, F. Hilbert, R. Lindqvist, M. Nauta, L. Peixe, G. Ru, M. Simmons, P. Skandamis, E. Suffredini, L. Castle, M. Crotta, K. Grob, M. R. Milana, A. Petersen, A. X. Roig Sagués, F. Vinagre Silva, E. Barthélémy, A. Christodoulidou, W. Messens and A. Allende, *EFSA J.*, 2022, **20**, e07128.

45 G. E. John, E. A. Okpo, J. Akpanke, C. U. Okoro, P. A. Omang and J. A. Lennox, *Afr. Health Sci.*, 2023, **23**, 758–763.