Handling the Impact of Uncertainties on Predicting the Quality Aspects of Doogh

Mohammad Amini Hoshiar[1], Sara Jafarian•[1], Ramezan Rezaian[2] , and Mahdi Sharifi Soltani[3]

Abstract-- **Support Vector Machines (SVMs) are valuable tools in the food industry due to their capability to handle complex, nonlinear relationships between variables, even with limited datasets, high-dimensional data, and noisy data. This makes SVMs well-suited for applications such as food quality and safety assessment, sensory evaluation, process optimization, and food authentication among others. Accordingly, a new approach is introduced to predict different features of a traditional yogurt drink, also called doogh. The proposed model combines the principles of Support Vector Regression with fuzzy logic to handle uncertainty and approximately complex relationships between inputs (retentate, xanthan, and shelf-life) and target variables including viscosity, syneresis, color values, and total acceptability. The implemented approach is particularly advantageous when dealing with problems where the relationships are not easily captured by traditional mathematical models due to their non-linearity or imprecision. Also, it mitigates the limitations of data availability. The predictive ability of the proposed model has been evaluated in terms of *MSE*, R^2, *RMSE*, and *MAE* when adding different noise levels. Additionally, the conditions necessary to attain optimized metric values have been found. At the optimum point, the viscosity, syneresis, L*, a*, b*, and total acceptability are 19.70 mPa.s, 11.30%, 97.04, - 1.43, 8.13, and 5.00, respectively. Besides, the findings indicate that samples containing 0.8% retentate, 0.4% xanthan, and a 31-day shelf-life exhibit the highest viscosity, while those with 0.6% retentate, 0.4% xanthan, and a 31-day shelf-life show the lowest syneresis. Moreover, samples with 0.7% retentate, 0.2% xanthan, and a 13-day shelf-life demonstrate the highest total acceptability.**

Index Terms- **Doogh, Fuzzy theory; Support Vector Regression (SVR); Syneresis; Uncertainty.**

I. INTRODUCTION

Doogh, a refreshing fermented drink from Iran, is made with yogurt, water, salt, and herbs [1]. This traditional product is also called "Aryan" in Turkey and "lassi" in India. This beverage is also produced in some regions of Eastern Europe and the Balkans [2].

Physical instability and syneresis are major problems of this product during shelf-life, which occurs due to the low viscosity of doogh [3]. Physical instability refers to the tendency of doogh ingredients to separate during storage. This separation can manifest as sedimentation (settling of solid particles (yogurt curds) at the bottom of the container), creaming (rising of lighter components (fat or whey) to the top), and flocculation (clumping of yogurt proteins). Syneresis refers to a particular form of physical instability where a liquid segregates from a gel. In the context of doogh, syneresis manifests as the expulsion of whey (the liquid component) from the yogurt gel. This results in an unappealing watery stratum on the surface and a denser yogurt layer at the bottom, which adversely affects the product's natural appearance and marketability [4]. One approach to prevent or delay the syneresis of doogh involves the use of hydrocolloid stabilizers [5,6]. Hydrocolloids enhance viscosity and establish a spatial barrier between colloidal particles, thereby stabilizing the fermented dairy beverage [7]. However, it's essential to determine the optimal hydrocolloid quantity. An insufficient amount may not effectively inhibit syneresis, while an excess can escalate overall costs and result in an overly thick or slimy texture, negatively affecting consumer preference. Consequently, numerous research studies have focused on establishing the optimal conditions for maximum stability with minimal stabilizer usage [8,9].

In doogh, xanthan gum can augment viscosity, enhance stability, and avert phase separation [12]. Rahmati et al. [2] investigated how adding xanthan gum, gelatin, or a mix of both impacted low-fat yogurt. They found that yogurts with a combination of 0.2% xanthan gum and 0.2% gelatin received the best scores for taste and texture (acceptability) and showed the least amount of liquid separation (syneresis). In another study, Gorji et al. [14] studied the influence of three types of Iranian tragacanth gum on the flow properties (rheology) and stability of non-fat yogurt. The results show that the yogurt with 0.3% Astragalus gossypinus tragacanth gum exhibited the greatest stability over a 30-day period.

During ultrafiltration, the casein proteins are retained by the membrane, while the lactose, minerals, and water pass through the membrane as permeate [17]. This results in a retentate that is enriched in casein proteins and has a higher viscosity than the original milk. The retentate can be used as a stabilizer in beverages, such as acidic beverages, to improve their texture and stability.

Despite complexities in quality, production, and logistics, the food industry is revolutionized by machine learning's efficient solutions, replacing traditional, time-consuming parameter acquisition [19]. Support Vector Machine (SVM)-based approaches are versatile machine learning algorithms capable of analyzing complex datasets and identifying patterns, making them invaluable for various applications in the food

1. Department of Food Science & Technology, Noor Branch, Islamic Azad University, Noor, Iran.
• Corresponding author Email: drsjafarian@yahoo.com
2. Department of Mathematics, Noor Branch, Islamic Azad University, Noor, Iran.
3. Department of Veterinary, Chalous Branch, Islamic Azad University, Chalous, Iran.

industry [20,21]. Also, SVM's robustness to noise and outliers in the data is essential in food engineering, where data can be noisy due to variability in raw materials, processing conditions, and measurement errors. Neural networks are the other well-known machine learning approaches that play a crucial role in the food industry, enhancing processes, safety, and overall quality. These approaches are capable of analyzing sensory data to assess food quality objectively, predict microbial growth, and estimate physical and chemical properties [22-24].

Uncertainty is a significant challenge in the food industry, and it can have various impacts on the quality, safety, and efficiency of food production. Some of the sources of uncertainty in the food industry include ingredient variability, environmental factors, process variability, equipment performance, human factors, supply chain dynamics, and consumer preferences. Managing and mitigating these uncertainties requires robust quality control measures, predictive modeling techniques, and continuous process improvement initiatives [25,26].

As mentioned above, machine learning-based approaches can have a significant impact on the food industry by improving the accuracy and efficiency of various processes, such as quality control, predictive modeling, and process optimization. However, methods proposed for predicting doogh characteristics (viscosity, syneresis, color, and total acceptability) lack uncertainty awareness, hindering accurate assessments [27,28]. The objective of this paper is to develop a predictive model for the quality attributes of doogh, taking into account process parameters and ingredient properties. This is achieved through the integration of support vector-based method with fuzzy theory. To ensure the effectiveness of the experimental conditions, the Taguchi method is utilized, enabling a methodical exploration of different factors. This approach optimizes the experimental arrangement by employing an appropriate Orthogonal Array (OA), which ensures thorough coverage of the parameter space while minimizing the number of experimental trials needed. Next, the experiments are conducted by the proposed model, and the values for syneresis, color values, and total acceptability are recorded.

The contribution of the present paper is as follows:

- The proposed model can predict values for syneresis, color, and total acceptability. Utilizing this model reduces testing costs and saves time.
- Due to data limitations, a support vector-based model integrated with the fuzzy theory is utilized during the training phase. This strategy proves beneficial when handling insufficient and imprecise data.

The remainder of the present paper is as follows. Section 2 reviews different applications of SVM in food processing. Section 3 presents materials and methods in detail. Section 4 discusses experiments. Finally, Section 5 describes the conclusions.

II. Related work

The attribute that makes SVM applicable in food engineering is its ability to handle complex, non-linear relationships between variables. In food engineering, there are often complex relationships between the physicochemical properties of food and their sensory, nutritional, and functional qualities [29,30]. SVM's kernel trick allows it to transform the input data into a higher-dimensional feature space, where it can capture these non-linear relationships and make accurate predictions. Moreover, SVM's ability to handle small sample sizes and high-dimensional data is also beneficial in food engineering, where data collection can be time-consuming and expensive [31]. SVM's sparsity property, which uses only a subset of the training data (support vectors) to define the decision boundary, makes it computationally efficient and scalable to large datasets [32]. Additionally, SVM's robustness to noise and outliers in the data is essential in food engineering, where data can be noisy due to variability in raw materials, processing conditions, and measurement errors [33]. SVM's maximal margin classification and soft margin classification properties make it less sensitive to noise and outliers, improving its generalization performance. Regarding these features, this section reviews different SVM applications in food processing, including rheology, fermentation, and shelf-life estimation.

Rheology is the study of the deformation and flow of matter under the influence of stresses (temperature, shear rate, time, moisture content, pressure, and density influence) and describes the physical properties of food materials [34]. Rheological measurements are essential for quality evaluation, viscosity profiling, engineering calculations like stability in extruders and mixers, process design, and standardizing dispersion stability of food products [35], e.g., in [36]. SVM enhanced plant-based cheese formulation through molecular docking and dynamic simulation of tocopherol and retinol complexes with zein, soy, and almond proteins.

Various studies suggest that SVM is an effective method for dynamic modeling and regulation of fermentation process parameters [40,41]. For instance, regarding chemical markers, SVMs were employed to detect abnormal wine fermentation [42,43]. In [44], some machine learning methods including SVM, random forest, partial least squares discriminant analysis, partial least squares regression, and neural networks contributed electronic nose for aroma quality characterization for Pixian broad bean paste fermentation. Besides, using SVM and convolutional neural networks [45], an online system was proposed for monitoring the degree of fermentation of oolong tea.

In addition, SVM can be employed for shelf-life evaluation of food products. In [48], an SVM strategy was introduced for representing the moisture of glutinous rice. In [49], an ensemble approach, which encompasses SVM, decision tree, random forest, and Gradient Boosting Machine (GBM), was proposed to develop a tomato ripeness and shelf-life prediction system based on defects and color intensity. In [50], an ensemble approach, which encompasses Linear Support Vector Regression (SVR), MLP, *k*NN, and decision tree algorithms, has been employed for gas sensor-based detection of salmon freshness and shelf-life.

SVMs have also been used for sensory evaluation applications such as modeling color parameters of cerasus humilis during ripening [51], predicting consumer liking scores of green tea beverages [52], and evaluating cheese texture during the ripening phase [36].

III. Materials and methods

This section comprises the following subsections: Subsection 3.1 details the methods involved in doogh preparation.

Subsection 3.2 explores the physicochemical analysis techniques used to characterize doogh. Subsection 3.3 presents the sensory evaluation process. Subsection 3.4 discusses the experimental design, including the parameters investigated, the levels considered, and the rationale behind the chosen experimental conditions. Subsection 3.5 presents the pre-processing step. Finally, Subsection 3.6 explains the proposed model in detail.

A. Preparation of doogh

Pasteurized milk containing 0.5% fat and retentate with 32.50% dry matter was obtained from Zarrin-Laban-Pars Company in Qazvin province in Iran. Xanthan gum was purchased from Sigma Aldrich. The starter culture was CH1, which includes *Lactobacillus delbrueckii* subspecies of *bulgaricus*. *Streptococcus thermophilus* was purchased from Christian Hansen in Denmark.

Samples of milk were heated in a viscobator until reaching a temperature of 43℃, and then inoculated with 2% starter culture. The retentate was added to the milk at the same temperature in three concentrations (0, 0.5, and 1% *w*/*v*). Fermentation ran at 42℃±1℃ until reaching a pH of 4.5.

Xanthan gum was dissolved in distilled water at three levels (0, 0.2, and 0.4% *w*/*v*) and 0.5% sodium chloride with a purity of 99% *w*/*v* was added for 2 minutes. Doogh samples were prepared by mixing yogurt and distilled water in a 50:50 ratio, and homogenized at 50-60 bars pressure using a homogenizer machine (Armfield, FT9 Homogenizer, England). The treatments were then pasteurized in a viscobator at 85℃ for 1 minute. A two-step cooling to 14 ˚C was applied, then the doogh samples were filled in 250 ml plastic bottles and further cooled to 5 ˚C. They were analyzed 24-h after the completion of fermentation (day 0) and on days 15 and 30.

B. Physicochemical analysis of doogh

Chemical analyses: Physicochemical properties, such as fat content, protein content, and pH, were measured using established methods presented in [53] to chemically analyze the yogurt samples.

Measurement of apparent viscosity: For the measurement of sample viscosity, a Brookfield DV-II+Pro viscometer (made in the USA) with the ULA spindle was used. The tests were conducted at a temperature of 20℃. The apparent viscosity of the samples was measured at a shear rate of 82.50 s^{-1}. This shear rate was close to the shear rate applied by the mouth on food products [54].

Color measurement: The color of the yogurt sample was measured using a Hunter Lab Flex Colorimeter (USA). The colorimeter was calibrated using a black and white tile, and then the samples were transferred to the device for testing. Color indices including L* (lightness), a* (greenness), and b* (yellowness) were measured and recorded based on the light reflectance [55].

C. Sensory evaluation

Fifteen trained panelists evaluated the taste, flavor, aroma, mouthfeel, and total acceptability of the product using a 5-point hedonic scoring method [56]. Each score ranged from 1 (very undesirable) to 5 (very desirable).

D. Experimental Design

As illustrated in Table I, the model focuses on three independent parameters: retentate, xanthan, and shelf-life, each with three distinct levels. To generate a training dataset for the model, experiments need to be performed involving these parameters, while recording viscosity, syneresis, color values, and total acceptability. The obtained dataset is presented in Table II. As can be seen, it contains both independent and target factors.

TABLE I
Independent Parameters and Levels

Level	**Retentate (*w*/*v* %)**	**Xanthan (*w*/*v* %)**	**Shelf-life (Day)**
1	0	0	1
2	0.5	0.2	15
3	1	0.4	30

The Taguchi method utilizes a special experimental layout called an orthogonal array. This array efficiently combines various levels of multiple factors into individual experiments. In this study, an L_{27} orthogonal array has been chosen to match the factors and their levels under investigation. Table II presents the experimental layout based on this chosen array. Each row in the table represents a unique trial condition, specifying the specific levels of each independent factor used in that experiment. Importantly, for each trial condition, the response variable (target factor) is measured three times to account for potential variations. Randomization of these trial runs further ensures the data collected is not influenced by the order of experiments.

E. Pre-processing

The input data should be normalized to be processed. The standardization method is as follows:

$$x_i = \frac{x_i^* - \min\limits_{1\le j\le n} x_j^*}{\max\limits_{1\le j\le n} x_j^* - \min\limits_{1\le j\le n} x_j^*} \quad (1)$$

Now, the new data $x_i \in [0,1]$ is dimensionless.

F. Fuzzy Support Vector Regression Model

The proposed model aims to predict the quality characteristics of doogh. It is important to note that data collected during food processing can be challenging to separate accurately and are frequently contaminated with noise and outliers, which impact the model's performance. To enhance the model, Fuzzy Support Vector Regression (FSVR) is employed, which offers control over the contribution of data points. FSVR assigns fuzzy membership values to each data point, allowing each point to influence the learning process differently, thereby reducing the impact of noise and outliers. Consequently, some data points are given more weight, while others lose their degree of importance. The main principles of FSVR are as follows:

Given the training sample $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}, y_i \in \mathcal{R}$, the regression model shaped as $f(x) = w^T x + b$ is the learning objective function. Moreover, the model parameters which make $f(x)$ as close as possible to y, w and b are to be determined by minimizing Eq. (2). SVR assumes that we can tolerate the maximum deviation of ε between $f(x)$ and y. That is to say, the loss is only calculated when the absolute value of the difference between $f(x)$ and y is greater than ε (see Eq. (3)). As shown in Fig. 1, this is equivalent to building a spacing band (i.e., the part sandwiched between the two dashed lines in the figure) with a width of ε centered on $f(x)$ (i.e., the solid line in the figure).

TABLE II
Dataset: Taguchi Experiments Based on L_{27} Orthogonal Array Design.

	Independent factors			Target factors					
#exp	**Retentate**	**Xanthan**	**Shelf-life**	**Visc.**	**Syn.**	**L***	**a***	**b***	**Total acceptability**
1	0	0	1	5.99	30.21	89.11	-1.55	6.12	3.89
2	0	0	1	6.81	30.05	89.24	-1.59	6.32	3.97
3	0	0	1	9.22	29.88	89.17	-1.53	6.27	4.03
4	0	0.2	15	11.17	24.34	91.52	-1.59	7.01	4.73
5	0	0.2	15	10.89	23.92	91.32	-1.67	7.08	4.59
6	0	0.2	15	11.42	24.5	91.41	-1.42	6.92	4.66
7	0	0.4	30	14.9	19.2	83.52	-1.82	7.14	3.29
8	0	0.4	30	14.01	18.79	84.02	-1.75	7.29	3.37
9	0	0.4	30	14.22	19.11	83.79	-1.89	7.19	3.32
10	0.5	0	15	10.27	20.55	94.11	-2.11	7.39	4.61
11	0.5	0	15	9.66	21.13	94.33	-2.16	7.44	4.49
12	0.5	0	15	9.93	20.96	94.05	-2.22	7.52	4.55
13	0.5	0.2	30	17.00	14.39	92.29	-1.95	7.13	4.29
14	0.5	0.2	30	16.23	14.46	92.17	-1.89	7.24	4.33
15	0.5	0.2	30	16.50	14.27	91.99	-1.93	7.06	4.37
16	0.5	0.4	1	14.75	16.37	87.11	-2.48	6.97	3.99
17	0.5	0.4	1	13.40	15.91	87.24	-2.59	6.89	3.96
18	0.5	0.4	1	18.51	16.11	87.04	-2.33	7.04	4.05
19	1	0	30	11.19	23.67	94.99	-2.48	7.78	4.21
20	1	0	30	8.14	24.01	95.16	-2.39	7.64	4.17
21	1	0	30	8.99	23.82	94.85	-2.51	7.69	4.25
22	1	0.2	1	13.11	17.22	92.93	-2.73	7.62	4.59
23	1	0.2	1	14.02	16.97	93.35	-2.9	7.49	4.48
24	1	0.2	1	13.67	17.03	93.17	-2.85	7.55	4.53
25	1	0.4	15	17.33	13.19	90.51	-2.93	7.95	4.01
26	1	0.4	15	17.60	12.78	90.39	-2.75	7.99	4.13
27	1	0.4	15	18.00	12.95	90.17	-2.82	8.03	3.92

Whereas if the training samples fall into this interval, they are considered to be correctly predicted.

$$R(f) = C \times \frac{1}{n}\sum_{i=1}^{n} L_{\varepsilon}(y_i, f_i) + \frac{1}{2}\|w\|^2 \quad (2)$$

$$L_{\varepsilon}(y_i, f_i) = \begin{cases} 0 & if\ |y_i - f_i| \le \varepsilon \\ |y_i - f_i| - \varepsilon & otherwise \end{cases} \quad (3)$$

where C and ε are user-defined parameters. Eq. (2) can be re-formalized as follows:

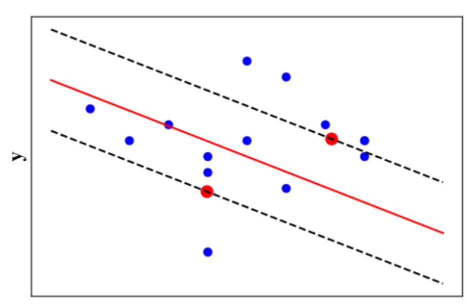

Fig. 1. A schematic for SVR

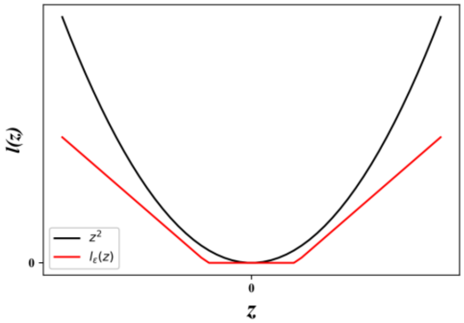

Fig. 2. The ε-insensitive loss function

$$\min_{w,b,\xi_i,\xi_i^*} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\mu_i\xi_i + \mu_i^*\xi_i^*) \tag{4}$$

Subject to $f(x_i) - y_i \le \varepsilon + \xi_i$

$$f(x_i) - y_i \le \varepsilon + \xi_i^*$$

$\xi_i\ ,\xi_i^* \ge 0, i = 1,2,\dots,n$

In this manuscript, the Gaussian membership function presented in (5) is used for calculating the membership function μ_i.

$$\mu_i = \sum_{j=1}^{\#\,of\ independent\ factors} \exp - \frac{(x_i^j - c_j)^2}{2{\sigma_j}^2} \tag{5}$$

where c is the center of the Gaussian function and is σ the standard deviation, which controls the width of the Gaussian curve. The solution of FSVR can be obtained by Lagrange multiplier method as follows:

$$\begin{aligned} \mathrm{L} = &\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\mu_i\xi_i + \mu_i^*\xi_i^*) \\ &- \sum_{i=1}^{n}(\eta_i\xi_i + \eta_i^*\xi_i^*) \\ &- \sum_{i=1}^{n}\alpha_i(\varepsilon + \xi_i - f(x_i) + y_i) \\ &- \sum_{i=1}^{n}\alpha_i^*(\varepsilon + \xi_i^* + f(x_i) - y_i) \end{aligned} \tag{6}$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0 \tag{7}$$

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^{n} x_i(\alpha_i - \alpha_i^*) = 0 \tag{8}$$

$$\frac{\partial L}{\partial \xi_i^*} = C\mu_i^* - \alpha_i^* - \eta_i^* = 0 \tag{9}$$

Apply these conditions into the Lagrangian Eq. (6), Eq. (4) can be transformed into

$$\begin{aligned} \min \frac{1}{2}\sum_{i,j=1}^{n}&(\alpha_i - \alpha_i^*)\,(\alpha_i - \alpha_i^*)K(x_i,x) \\ &+ \varepsilon\sum_{i=1}^{n}(\alpha_i + \alpha_i^*) \\ &- \sum_{i=1}^{n} y_i(\alpha_i^* - \alpha_i) \end{aligned} \tag{10}$$

Subject to $\sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0$

$0 \le \alpha_i, \alpha_i^* \le C\mu_i, i = 1,2,\dots,n$

The KKT conditions are defined as

$$\begin{aligned} &\alpha_i(\varepsilon + \xi_i + f(x_i) - y_i + b) \\ &\quad = 0 \\ &\alpha_i^*(\varepsilon + \xi_i - f(x_i) + y_i - b) \\ &\quad = 0 \end{aligned} \tag{11}$$

and

$$\begin{aligned} &(C\mu_i - \alpha_i) - \xi_i = 0 \\ &(C\mu_i^* - \alpha_i^*) - \eta_i^* = 0 \end{aligned} \tag{12}$$

So,

$$\begin{aligned} &b = y_i - f(x_i) - \xi_i \text{ for } \alpha_i \in (0, C\mu_i) \\ &b = y_i - f(x_i) + \xi_i \text{ for } \alpha_i^* \in (0, C\mu_i) \end{aligned} \tag{13}$$

Finally, the expression of the separation hyperplane becomes

$$f(x) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)K(x_i,x) + b \tag{14}$$

where $K(x_i,x) = \phi(x_i)^T\phi(x)$ is expressed as the Gaussian kernel function where $\phi(.)$ is the mapping in a space where dot-product can be expressed using the kernel property.

Here, adjusting the C (for penalty parameter of error term) is done using grid search. Grid search optimization algorithm is a kind of exhaustive search algorithm which optimizes the parameters by cross validation, and then obtains the optimal learning algorithm. The range of C is set $\{1,10,100,1000\}$.

IV. Experimental Results

In this paper four error metrics including *MSE*, R^2, *RMSE*, and *MAE* are used for evaluating the performance of the proposed model in predicting the viscosity, syneresis, color values, and total acceptability. According to the findings presented in Table III and Fig. 3, the proposed model gains high R^2 value for syneresis, for L* has a high R^2 value, for a* has a moderately low R^2 value but low error metrics. Besides, FSVR for b* has the best performance with a very high R^2 value and very low *MSE, RMSE,* and *MAE*. FSVR for Total acceptability obtains a high R^2 value. The proposed model for b* seems to perform the best, while for a* has room for improvement in terms of variance explanation (R^2).

Furthermore, the employed model has determined the optimal values for viscosity, syneresis, L*, a*, b*, and total acceptability. Moreover, it has been identified that these optimal parameters correspond to specific conditions (retentate, xanthan, and shelf-life). Table IV presents these findings in detail.

TABLE III

The Performance Evaluation of FSVR. The Ranks are Shown in Parentheses. The Best Results are Shown in Bold.

Metric	**Visc.**	**Syn.**	**L***	**a***	**b***	**Total acceptability**
MSE	0.079 (3)	0.088 (4)	0.097 (5)	0.077 (2)	**0.069 (1)**	0.099 (6)
R^2	0.890 (2)	**0.989 (1)**	**0.989 (1)**	0.668 (4)	0.974 (3)	**0.899 (1)**
$RMSE$	0.281 (3)	0.296 (4)	0.312 (5)	0.278 (2)	**0.263 (1)**	0.315 (6)
MAE	0.230 (3)	0.207 (2)	0.263 (5)	0.239 (4)	**0.172 (1)**	0.294 (6)
Average rank	2.750	2.750	4.000	3.000	**1.500**	4.750

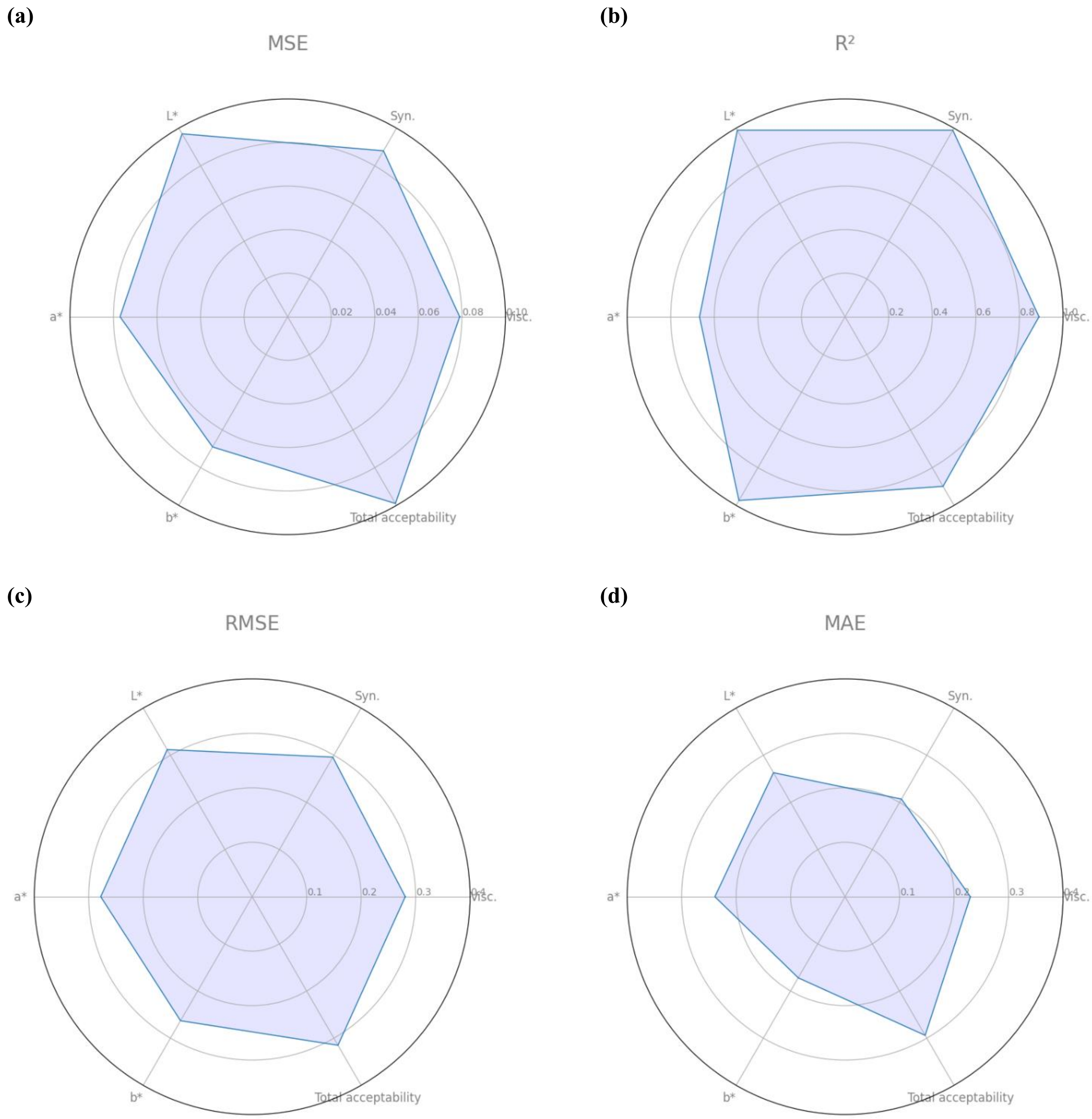

Fig. 3. The performance of the FSVR in terms of different error metrics. (a) MSE, (b) R^2, (c) $RMSE$, and (d) MAE.

TABLE IV
Predicted Optimized Parameters and Their Values Using FSVR.

Parameter	Predicted retentate (w/v %)	Predicted xanthan (w/v %)	Predicted shelf-life (Day)	Predicted value for parameter
Visc.	0.8	0.4	31	19.70
Syn.	0.6	0.4	31	11.30
L*	1	0.1	18	97.04
a*	0	0.1	31	-1.43
b*	1	0	15	8.13
Total acceptability	0.7	0.2	13	5.00

The findings indicate that the samples containing 0.8% retentate, 0.4% xanthan, and a 31-day shelf-life exhibit the highest viscosity, while those with 0.6% retentate, 0.4% xanthan, and a 31-day shelf-life show the lowest syneresis. Moreover, samples with 0.7% retentate, 0.2% xanthan, and a 13-day shelf-life demonstrate the highest total acceptability.

Referring to TableIV, the texture of doogh, particularly influenced by 0.7% retentate, aligns with panelists' preferences, potentially enhancing total acceptability, in accordance with Rahmati et al. [2]. Moreover, xanthan likely contributes to syneresis. Even though xanthan's direct impact on total acceptability has not been explicitly mentioned [57], the results show it likely contributes to syneresis. Xanthan also influences doogh's viscosity, leading to a thicker consistency and slower phase separation compared to treatment samples. According to [58], doogh samples containing 0.25% ginger extract and 0.5% gum arabic gained the highest total acceptability score compared to other samples but in the present study, 0.2% xanthan contributes to a pleasant total acceptability.

A. Noise study

To assess the robustness of the proposed model to noisy data, random noise from a normal distribution is added to the dataset at five noise levels: $\{5\%, 10\%, 15\%, 20\%, 25\%\}$. The model's performance is evaluated on these noisy datasets. The results are summarized in Table V and Fig. 4.

TABLE V
The Performance Evaluation of FSVR in the Presence of Different Levels of Noise.

Metric	Noise level (%)	Visc.	Syn.	L*	a*	b*	Total acceptability
MSE	0	0.079	0.088	0.097	0.077	0.069	0.099
	5	0.087	0.094	0.104	0.08	0.078	0.103
	10	0.087	0.098	0.111	0.08	0.087	0.106
	15	0.095	0.102	0.112	0.085	0.095	0.110
	20	0.103	0.111	0.116	0.091	0.098	0.114
	25	0.108	0.116	0.119	0.095	0.103	0.118
R^2	0	0.890	0.989	0.989	0.668	0.974	0.899
	5	0.872	0.969	0.969	0.655	0.955	0.881
	10	0.855	0.950	0.950	0.642	0.936	0.863
	15	0.838	0.931	0.931	0.629	0.917	0.846
	20	0.821	0.912	0.912	0.616	0.899	0.829
	25	0.805	0.894	0.894	0.604	0.881	0.812
$RMSE$	0	0.281	0.297	0.311	0.277	0.263	0.315
	5	0.295	0.307	0.322	0.283	0.279	0.321
	10	0.295	0.313	0.333	0.283	0.295	0.326
	15	0.308	0.319	0.335	0.292	0.308	0.332
	20	0.321	0.333	0.341	0.302	0.313	0.338
	25	0.329	0.341	0.345	0.308	0.321	0.344
MAE	0	0.230	0.207	0.263	0.239	0.172	0.294
	5	0.322	0.343	0.344	0.309	0.316	0.339
	10	0.323	0.353	0.345	0.309	0.318	0.345
	15	0.332	0.357	0.345	0.314	0.322	0.351
	20	0.334	0.358	0.346	0.319	0.329	0.36
	25	0.337	0.366	0.35	0.324	0.334	0.364

As can be shown, the performance of the FSVR degrades as the level of noise in the data increases. This is evident from the increasing MSE, $RMSE$, and MAE values and the decreasing R^2 values across all metrics. Table VI quantifies the magnitude of these changes. Notably, b* and syneresis are more sensitive to noise, showing a more significant performance decline compared to a* and Total acceptability

.

TABLE VI
The Magnitude of Changes for Metrics. The Ranks are Shown in Parentheses. The Best Results are Shown in Bold.

Metric	Visc.	Syn.	L*	a*	b*	Total acceptability
MSE	0.029 (5)	0.028 (4)	0.022 (3)	**0.018 (1)**	0.034 (6)	0.019 (2)
R^2	0.069 (2)	0.077 (5)	0.077 (5)	**0.052 (1)**	0.075 (4)	0.070 (3)
$RMSE$	0.040 (5)	0.036 (4)	0.030 (3)	0.025 (2)	0.050 (6)	**0.023 (1)**
MAE	0.107 (4)	0.159 (5)	0.087 (3)	0.085 (2)	0.162 (6)	**0.070 (1)**
Average rank	4.000	4.500	3.500	**1.500**	5.500	1.750

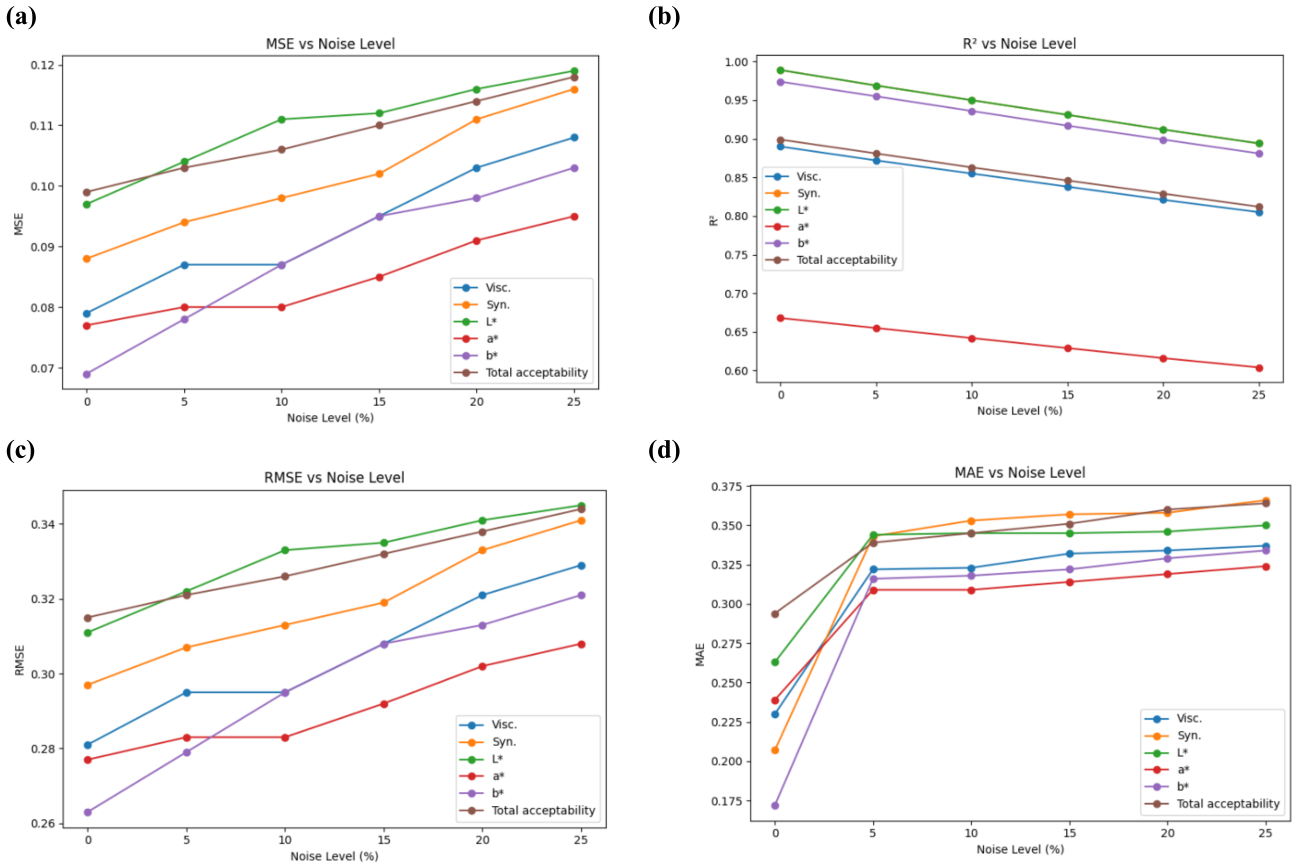

Fig. 4. The performance of the employed FSVR in terms of different error metrics when different levels of noisy data are added. (a) MSE, (b) R^2, (c) $RMSE$, and (d) MAE.

V. Conclusion

This study leverages the multifaceted role of Support Vectors (SVs) to improve quality control, optimize parameters, and enhance overall efficiency in the prediction of various rheological, physicochemical, and sensorial properties of doogh. To address the inevitable presence of uncertainty in measured data, Fuzzy Support Vector Regression (FSVR) is utilized, which assigns membership degrees to each data point, allowing for individual weights of participation in the model. The proposed model demonstrates superior performance in predicting crucial properties related to the product's rheology, physicochemistry, and sensory attributes, thereby reducing the need for time-consuming and expensive experimental procedures. The optimized parameters, including viscosity, syneresis, L*, a*, b*, and total acceptability, have been calculated to provide insights into the optimal conditions for achieving desired characteristics in doogh. The results indicate that the optimum viscosity is 19.70 mPa.s, syneresis is 11.30%, and the values for L*, a*, b*, and total acceptability are 97.04, -1.43, 8.13, and 5.00, respectively. Furthermore, the findings suggest that the samples containing 0.8% retentate, 0.4% xanthan, and a 31-day shelf-life exhibit the highest viscosity, while those with 0.6% retentate, 0.4% xanthan, and a 31-day shelf-life show the lowest syneresis. Additionally, samples with 0.7% retentate, 0.2% xanthan, and a 13-day shelf-life demonstrate the highest total acceptability. These findings have practical implications for food product formulation and quality control, enabling manufacturers to optimize their formulations, reduce testing costs, and enhance product stability by leveraging this predictive model.

REFERENCES

[1] Azarikia F, Abbasi S (2010) On the stabilization mechanism of Doogh (Iranian yoghurt drink) by gum tragacanth. Food Hydrocolloids 24 (4):358-363

[2] Rahmati F, Mahjoorian A, Fazeli F, Ranjbar S (2023) Investigation of rheological, physicochemical, and sensorial properties of traditional low-fat Doogh formulated. Food Science & Nutrition 11 (11):7218-7228

[3] Turkmen N, Akal C, Özer B (2019) Probiotic dairy-based beverages: A review. Journal of Functional Foods 53:62-75

[4] Tahsiri Z, Niakousari M, Khoshnoudi-Nia S, Hosseini SMH (2017) Sensory evaluation of selected formulated milk barberry drinks using the fuzzy approach. Food science & nutrition 5 (3):739-749

[5] Mani-López E, Palou E, López-Malo A (2014) Probiotic viability and storage stability of yogurts and fermented milks prepared with several mixtures of lactic acid bacteria. Journal of Dairy Science 97 (5):2578-2590. doi:https://doi.org/10.3168/jds.2013-7551

[6] DeBruyne AN, Hekmat S (2024) The effects of fortification of yogurt with various functional flours on survival and growth of probiotic bacteria and sensory properties of the yogurt. Nutrition & Food Science 54 (3):597-612. doi:10.1108/NFS-11-2023-0257

[7] Beirami-Serizkani F, Hojjati M, Jooyandeh H (2021) The effect of microbial transglutaminase enzyme and Persian gum on the characteristics of traditional kefir drink. International Dairy Journal 112:104843

[8] Sworn G (2021) Xanthan gum. Handbook of hydrocolloids. Elsevier, pp 833-853

[9] Karagözlü C, Kınık Ö, Sibel Akalın A, Kesenkaş H, Dinkçi N (2024) Characterisation of probiotic fermented milks produced from goat milk containing low or high somatic cell counts. Mljekarstvo: časopis za unaprjeđenje proizvodnje i prerade mlijeka 74 (2):106-115

[10] Khoobbakht F, Khorshidi S, Bahmanyar F, Hosseini SM, Aminikhah N, Farhoodi M, Mirmoghtadaie L (2024) Modification of mechanical, rheological and structural properties of agar hydrogel using xanthan and locust bean gum. Food Hydrocolloids 147:109411

[11] Bak JH, Yoo B (2018) Intrinsic viscosity of binary gum mixtures with xanthan gum and guar gum: Effect of NaCl, sucrose, and pH. International journal of biological macromolecules 111:77-81

[12] Cheng Y, Wang B, Lv W, Zhong Y, Li G (2024) Effect of xanthan gum on physicochemical properties and 3D printability of emulsion-filled starch gels. Food Hydrocolloids 149:109613. doi:https://doi.org/10.1016/j.foodhyd.2023.109613

[13] Zhang J, Wang G, Liang Q, Cai W, Zhang Q (2019) Rheological and microstructural properties of gelatin B/tara gum hydrogels: Effect of protein/polysaccharide ratio, pH and salt addition. LWT 103:108-115. doi:https://doi.org/10.1016/j.lwt.2018.12.080

[14] Gorji EG, Mohammadifar MA, Ezzatpanah H (2011) Influence of gum tragacanth, Astragalus gossypinus, addition on stability of nonfat Doogh, an Iranian fermented milk drink. International Journal of Dairy Technology 64 (2):262-268

[15] Reig M, Vecino X, Cortina JL (2021) Use of membrane technologies in dairy industry: An overview. Foods 10 (11):2768

[16] Brazuelo A, Suarez E, Riera FA, Alvarez R, Igesias JR, Granda J (1995) Protein-enriched yoghurt by ultrafiltration of skim-milk. Journal of the Science of Food and Agriculture 69 (3):283-290

[17] Singh R, Rathod G, Meletharayil GH, Kapoor R, Sankarlal VM, Amamcharla JK (2022) Invited review: Shelf-stable dairy protein beverages—Scientific and technological aspects. Journal of Dairy Science 105 (12):9327-9346

[18] Amatayakul T, Sherkat F, Shah NP (2006) Physical characteristics of set yoghurt made with altered casein to whey protein ratios and EPS-producing starter cultures at 9 and 14% total solids. Food hydrocolloids 20 (2-3):314-324

[19] Tajkarimi M (2020) Food Safety and Quality Data Management Using Artificial Intelligence. Food Protection Trends 40 (6):464-467

[20] Ni X, Jiang Y, Zhang Y, Zhou Y, Zhao Y, Guo F, Wang H (2023) Identification of liquid milk adulteration using Raman spectroscopy combined with lactose indexed screening and support vector machine. International Dairy Journal 146:105751. doi:https://doi.org/10.1016/j.idairyj.2023.105751

[21] Amsaraj R, Ambade ND, Mutturi S (2021) Variable selection coupled to PLS2, ANN and SVM for simultaneous detection of multiple adulterants in milk using spectral data. International Dairy Journal 123:105172. doi:https://doi.org/10.1016/j.idairyj.2021.105172

[22] Lazim MA, Suriani M (2009) Sensory evaluation of the selected coffee products using fuzzy approach. World Academy of Science, Engineering and Technoly International Journal of Mathematical and Computational Sciences 50 (2):133-136

[23] Vishwakarma S, Mandliya S, Dalbhagat CG, Singh PK, Mishra HN (2023) Rheological modeling of marjoram fortified rice dough: Empirical and machine learning approach. Journal of Food Process Engineering 46 (8):e14387

[24] Zhou Z, Tian D, Yang Y, Cui H, Li Y, Ren S, Han T, Gao Z (2024) Machine learning assisted biosensing technology: An emerging powerful tool for improving the intelligence of food safety detection. Current Research in Food Science:100679

[25] Martins JA, Azevedo AM, de Almeida AC, da Silva LCR, Fernandes ACG, Valadares NR, Aspiazu I (2022) Fuzzy logic is a powerful tool for the automation of milk classification. Acta Scientiarum Technology

[26] Meena GS, Gupta VK, Khetra Y, Raghu HV, Parmar PT (2015) Fuzzy logic (Similarity Analysis) modelling for sensory evaluation of the market samples of Kheer Mohan. Indian J Dairy Sci 68 (4):326-333

[27] Alves V, dos Santos JM, Pinto E, Ferreira IM, Lima VA, Felsner ML (2024) Digital image processing combined with machine learning: A new strategy for brown sugar classification. Microchemical Journal 196:109604

[28] Abass T, Itua EO, Bature T, Eruaga MA (2024) Concept paper: Innovative approaches to food quality control: AI and machine learning for predictive analysis. World Journal of Advanced Research and Reviews 21 (3):823-828

[29] Zhang X, Li H, Mu W, Gecevska V, Zhang X, Feng J (2022) Sensory evaluation and prediction of bulk wine by physicochemical indicators based on PCA-PSO-LSSVM method. Journal of Food Processing and Preservation 46 (3):e16343

[30] Issaad FZ, Abdessemed A, Bouhedjar K, Bouyahmed H, Derdour M, Ouffroukh K, Fellak A, Dems MAS, Chihoub S, Bechlem R (2024) Classification of Algerian olive oils: Physicochemical properties, polyphenols and fatty acid composition combined with machine learning models. Journal of Food Composition and Analysis 125:105812

[31] 31. Sun J, Yang F, Cheng J, Wang S, Fu L (2024) Nondestructive identification of soybean protein in minced chicken meat based on hyperspectral imaging and VGG16-SVM. Journal of Food Composition and Analysis 125:105713

[32] Singh R (2024) Implementation of User Rating Classification for Amazon FoodReview Dataset Using SVM and LSTM. International Journal of Intelligent Systems and Applications in Engineering

[33] Wang Y, Gu H-W, Yin X-L, Geng T, Long W, Fu H, She Y (2024) Deep leaning in food safety and authenticity detection: An integrative review and future prospects. Trends in Food Science & Technology:104396

[34] Ahmad I, Benjamin T-A (2023) Chapter 8 - Application of artificial intelligence and machine learning to food rheology. In: Ahmed J, Basu S (eds) Advances in Food Rheology and Its Applications (Second Edition). Woodhead Publishing, pp 201-219. doi:https://doi.org/10.1016/B978-0-12-823983-4.00004-2

[35] Boger DV, Tiu C (1974) Rheological properties of food products and their use in the design of flow systems. Food technology in Australia

[36] Yakoubi S (2024) Enhancing plant-based cheese formulation through molecular docking and dynamic simulation of tocopherol and retinol complexes with zein, soy and almond proteins via SVM-machine learning integration. Food Chemistry 452:139520. doi:https://doi.org/10.1016/j.foodchem.2024.139520

[37] Croonenberghs AP, Bongaerts D, Bouchez A, De Roos J, De Vuyst L (2024) Fruit beers, beers with or without a co-fermentation step with fruits. Current Opinion in Biotechnology 86:103081

[38] Sheikh H, Anand GG, Shivanna GB (2024) Fermentation: A Potential Strategy for Microbial Metabolite Production.

[39] Ma X, Wang Y, Liu Y, Li X, Wang F, Huang Y, Shi P, Brennan CS, Wang M (2024) Mechanisms and factors influencing the ability of lactic acid bacteria on reducing biogenic amines in fermented food: A mini-review. LWT:115890

[40] Lončar B, Pezo L, Iličić M, Kanurić K, Vukić D, Degenek J, Vukić V (2024) Modeling and Optimization of Herb-Fortified Fresh Kombucha Cheese: An Artificial Neural Network Approach for Enhancing Quality Characteristics. Foods 13 (4):548

[41] Yu H, Liu S, Qin H, Zhou Z, Zhao H, Zhang S, Mao J (2024) Artificial intelligence-based approaches for traditional fermented alcoholic beverages' development: Review and prospect. Critical Reviews in Food Science and Nutrition 64 (10):2879-2889

[42] Urtubia A, León R, Vargas M (2021) Identification of chemical markers to detect abnormal wine fermentation using support vector machines. Computers & Chemical Engineering 145:107158. doi:https://doi.org/10.1016/j.compchemeng.2020.107158

[43] Hernández G, León R, Urtubia A (2016) Detection of abnormal processes of wine fermentation by support vector machines. Cluster Computing 19:1219-1225

[44] Xu M, Wang X, Xu Z, Wang Y, Jia P, Ding W, Dong S, Liu P (2024) Aroma quality characterization for Pixian broad bean paste fermentation by electronic nose combined with machine learning methods. Journal of Food Measurement and Characterization 18 (5):3363-3377

[45] Zheng P, Solomon Adade SY, Rong Y, Zhao S, Han Z, Gong Y, Chen X, Yu J, Huang C, Lin H (2024) Online System for Monitoring the Degree of Fermentation of Oolong Tea Using Integrated Visible–Near-Infrared Spectroscopy and Image-Processing Technologies. Foods 13 (11). doi:10.3390/foods13111708

[46] Nikravan L, Zamanpour S, Noori SMA (2024) Postbiotics: an innovative approach to increase shelf life and quality of foods. Nutrition & Food Science 54 (1):192-206

[47] Phimolsiripol Y, Suppakul P (2016) Techniques in shelf life evaluation of food products.

[48] Jimoh KA, Hashim N, Shamsudin R, Che Man H, Jahari M (2024) Optimization of computational intelligence approach for the prediction of glutinous rice dehydration. Journal of the Science of Food and Agriculture 104 (10):6208-6220

[49] Goyal K, Kumar P, Verma K (2024) Tomato ripeness and shelf-life prediction system using machine learning. Journal of Food Measurement and Characterization 18 (4):2715-2730

[50] Liang B, Li X, Yang M, Zhang Z, Ren J (2024) Stacking ensemble learning for gas sensor-based detection of salmon freshness and shelf life. Journal of Food Process Engineering 47 (3):e14593

[51] Wang B, Yang H, Li L, Zhang S (2024) Non-Destructive Detection of Cerasus Humilis Fruit Quality by Hyperspectral Imaging Combined with Chemometric Method. Horticulturae 10 (5):519

[52] Bagnulo E, Strocchi G, Bicchi C, Liberto E (2024) Industrial food quality and consumer choice: Artificial intelligence-based tools in the chemistry of sensory notes in comfort foods (coffee, cocoa and tea). Trends in Food Science & Technology:104415

[53] Gharibzahedi SMT, Chronakis IS (2018) Crosslinking of milk proteins by microbial transglutaminase: Utilization in functional yogurt products. Food chemistry 245:620-632

[54] Afonso IM, Maia JM (1999) Rheological monitoring of structure evolution and development in stirred yogurt. Journal of Food Engineering 42 (4):183-190

[55] Ghosi Hoojaghan S, Sedaghati M, Mooraki N (2022) Characterization of Iranian Doogh enriched with gum Tragacanth and fennel extract (Foeniculum Vulgare). Journal of Agricultural Science and Technology 24 (6):1345-1356

[56] Janhøj T, Frøst MB, Ipsen R (2008) Sensory and rheological characterization of acidified milk drinks. Food Hydrocolloids 22 (5):798-806

[57] Khodashenas M, Jouki M (2020) Optimization of stabilized probiotic Doogh formulation by edible gums and response surface methodology: assessment of stability, viability and organoleptic attributes. Journal of Food Science and Technology 57:3201-3210

[58] Haji Ghafarloo M, Jouki M, Tabari M (2020) Production and characterization of synbiotic Doogh, a yogurt-based Iranian drink by gum Arabic, ginger extract and B. bifidum. Journal of food science and technology 57:1158-1166