Enhancing tobacco leaf similarity research through multi-modal feature fusion

Peiyao Gong, *Yuxuan* Dong, and *Kaihu* Hou*

Kunming University of Science and Technology, College of mechanical and electrical engineering, 650000 Kunming Yunnan, China

Abstract. The study proposed a method for identifying tobacco leaves with similar quality to stored leaves. It calculated leaf similarity using multi-modal fusion. It extracted near-infrared spectral features with a 1D-CNN network and image features with the correlation function in OpenCV. Then, it applied the attention mechanism feature fusion method to combine these features. The method calculated the similarity between tobacco leaves using Euclidean distance, which allows for the identification of replacement leaves most similar to the target leaves. This approach effectively fulfills the objective of sustaining the cigarette leaf group formula during maintenance by addressing the requirement to identify comparable tobacco leaves in instances where a specific raw material is lacking.

1 Introduction

As tobacco is a unique agricultural product that is heavily influenced by environmental and climatic conditions in its growing regions, the quality and yield of different grades of tobacco leaves can vary significantly and are challenging to manage. In the production process, if there is a shortage or fluctuation in the quality of a particular tobacco leaf in the formula, it must be replaced with another leaf of similar quality. Traditionally, this search for alternative tobacco leaves relied on physical and chemical indicators, expert experience, and sensory evaluations from smoking tests. However, this approach is time-consuming, costly, subjective, and unreliable in terms of timeliness and quality [1].

To address these challenges, this study proposes a novel method for calculating tobacco leaf similarity based on multi-modal fusion. This digital approach streamlines the search for comparable tobacco leaves, reducing costs, increasing efficiency, and providing a more effective means for maintaining leaf group formulations.

2 Experimental equipment and materials

In this experiment, 25 high-frequency key-grade tobacco leaves redried by K326 in Yunnan Province in 2023 were used. Through Nicolet's Antaris II near-infrared spectrometer, 3021 near-infrared spectral samples of tobacco leaves were collected, averaging 64 scanned interferograms at 8cm^{-1} resolution. The near infrared range is from $3800cm^{-1}$ to

* Corresponding author: 2803342872@qq.com

$10000cm^{-1}$.The samples were divided into an 8:2 ratio for the training and test sets, respectively. Details on sample quantities and production areas can be found in Table 1, with near-infrared spectra of the 25 grades of tobacco leaves showcased in Figure 1.

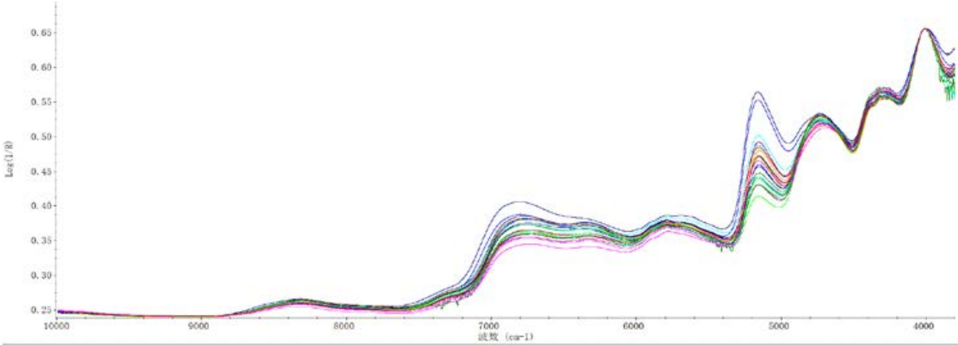
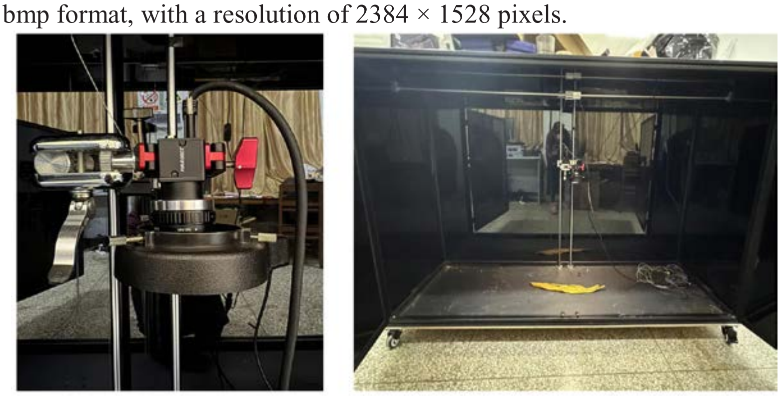


Fig. 1. Near infrared spectra of tobacco leaves.

Table 1. Near-infrared spectral data sets of tobacco leaf samples from different producing areas.

Producing area	Sample size		
	Training sets	Testing set	Total
A	490	123	613
B	309	77	386
C	360	90	450
D	503	126	629
E	430	107	537
F	325	81	406

To ensure a consistent light source, pixel quality, and image environment for acquiring tobacco leaf images, a specialized platform was constructed. The platform, depicted in Figure 2, features a dark box that isolates natural light, housing a standardized light source for uniform lighting conditions during image acquisition. The Hikvision MVS client was utilized to capture tobacco leaf images by connecting to a computer. The image acquisition system includes a Haikang robot industrial array camera MV-CA050-12UC , Haikang robot industrial lens MVL-KF0618M-12MPE, dark box 2M×0.7M×1.2M, and ring light source OK65 . Different grades of tobacco leaves were sequentially placed in the dark box for image capture, ensuring they are centered in the image and within the camera's field of view. Images were stored in bmp format, with a resolution of 2384 × 1528 pixels.



（a） Industry camera （b） Image acquisition box

Fig. 2. Tobacco leaf collection equipment.

3 Near-infrared spectral feature extraction of flue-cured tobacco based on one-dimensional convolutional neural network

3.1 1D-CNN network to extract near-infrared spectral features

This study employed a 1D-CNN network to analyze the near-infrared spectrum of flue-cured tobacco. Unlike traditional CNNs, the 1D-CNN modifies the convolution kernel shape to 1 × n, making it better suited for one-dimensional data like spectra. The specific convolution process is shown in Fig.3, As illustrated in Fig. 4, the model comprises convolution (blue), fully connected (orange), and output layers (green) [2].

Initially, 1609 spectral sequences were fed into multiple convolutional layers, which extract spectral features. These features were then processed through pooling to reduce dimensionality, followed by Batch Normalization for improved model generalization. Subsequently, a fully connected layer mapped features to output categories [3]. Finally, after removing the fully connected layer, features were outputted from the pooling layer, resulting in a final feature dimension of 256 [4].

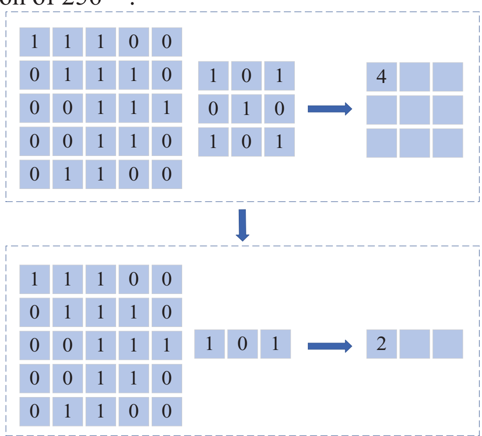


Fig. 3. Convolution process.

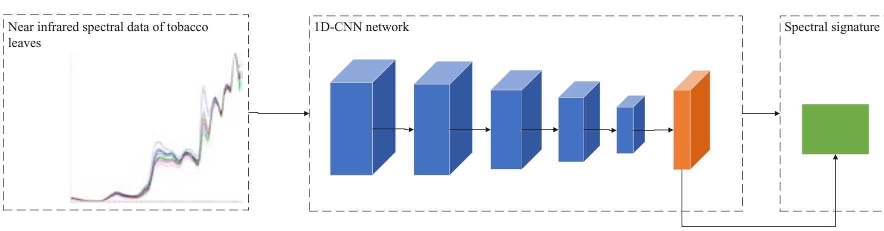


Fig. 4. 1D-CNN network model framework diagram.

The 1D-CNN model uses the ReLU (Rectified Linear Unit) function as the activation function, and its expression is :

$$f(x) = ReLU(x) = \max(0, x) = \begin{cases} 0, x \le 0 \\ x, x > 0 \end{cases} \tag{1}$$

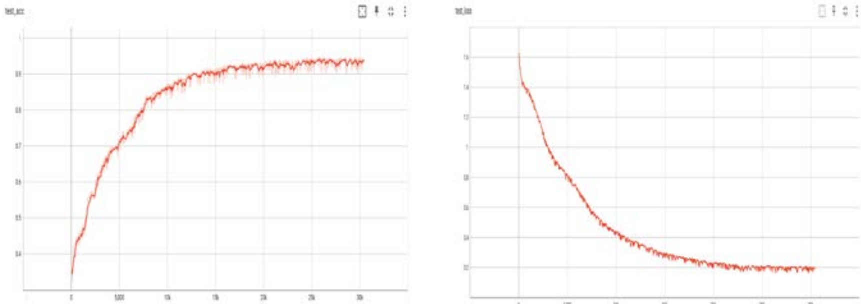
The range of ReLU function is$[0, +\infty]$, the input eigenvalue output less than 0 is 0, and the non-negative input remains unchanged, which makes the input sparse and reduces the computational cost.

In order to better deal with class imbalance and difficult samples in the training process and improve the performance and robustness of the model, the Focal Loss (FL) function was introduced as the loss function. By adjusting the weight of each sample, the model paid more attention to difficult samples in the training process, thereby improving the performance of the model on difficult samples. Its expression is :

$$FL = -a_t \left(1 - p_t\right)^{\gamma} \log\left(p_t\right) \tag{2}$$

In the equation, p_t denotes the prediction probability obtained from the model output after softmax processing. γ is a hyperparameter that adjusts the weight assigned to challenging samples. A small p_t leads to a larger $(1-p_t)^{\gamma}$, increasing the loss value and focusing attention on the sample. Conversely, a large p_t results in a smaller $(1-p_t)^{\gamma}$, reducing the loss and minimizing the impact on easier samples [5].

Analysis of the 1D-CNN network's accuracy and loss, as shown in Figure 5, revealed a trend over 30,000 training iterations. The graph's x-axis indicated the number of iterations, while (a) the y-axis displays accuracy and (b) showcases loss. The model's accuracy initially rose sharply before gradually stabilizing, whereas the loss declines rapidly at the start of training, eventually tapering off upon convergence. Training accuracy peaks at 0.9493 after 28,534 iterations, corresponding to a loss of 0.21.



（a）1D-CNN network accuracy change map （b）1D-CNN network loss value change diagram

Fig. 5. 1D-CNN network training map.

3.2 Correlation analysis between near infrared spectroscopy and sensory evaluation characteristics

The 256 dimensions of spectral features extracted by 1D-CNN were found to contain numerous redundancies and low correlation with the taste of tobacco leaves. In an effort to replace traditional leaf group formula maintenance with sensory evaluation, the sensory information was expressed through near-infrared spectral features [6]. Correlation analysis was conducted between 256 near-infrared spectral features and sensory evaluation features to identify significant spectral characteristics related to aroma quality, quantity, offensive odor, irritation, and aftertaste. Eight characteristics (serial numbers 41, 42, 56, 68, 99, 149, 234, and 235) were selected based on correlation coefficients ≥0.5, as shown in Figure 6.

	41	42	56	68	99	149	234	235
1	0.22	0.47	0.5	0.14	0.6	0.21	0.32	0.21
2	0.31	0.51	0.27	0.059	0.44	0.25	0.3	0.26
3	0.39	0.61	-0.053	0.5	0.19	0.51	0.51	0.5
4	0.3	0.53	0.22	0.29	0.43	0.55	0.59	0.37
5	0.5	0.55	0.028	0.41	0.4	0.46	0.53	0.39

Fig. 6. Correlation analysis between near infrared spectral characteristics and sensory evaluation of tobacco leaves.

4 Feature extraction of flue-cured tobacco image based on machine vision

4.1 Image characteristic extraction

During the image capture or transmission process, noise may be introduced, leading to a decrease in image quality and causing pixel values to diverge from their true values [7]. To address this issue, a Gaussian filter is employed for image denoising. The principle formula of the Gaussian filter is:

$$G(x,y)=\frac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{3}$$

In the formula, (x, y) signifies the coordinates within the image, while σ represents the standard deviation of the Gaussian function. This parameter dictates the breadth of the Gaussian function, thereby influencing the extent of filtering applied to the image.

707 images were extracted at various levels, using the correlation function in the OpenCV library to obtain 31 features such as shape, color, and texture of the tobacco leaf image. The length and width features were extracted using approximations, with the function cv2.minAreaRect() utilized to create the minimum circumscribed rectangle [8]. The length and width of this rectangle were then utilized as the respective features for the tobacco leaf.

To represent the color of tobacco leaves in RGB channels, the image values in the tobacco leaf image are converted into color moments. The first-order moment of color represents the average values of the R, G, and B components, effectively capturing the primary color distribution of the tobacco leaves [9]. The calculation formulas for each component are as follows, utilizing the OpenCV toolkit:

$$R=\sum_{i=1}^{n}\frac{r_i}{N} \qquad G=\sum_{i=1}^{n}\frac{g_i}{N} \qquad B=\sum_{i=1}^{n}\frac{b_i}{N} \tag{4}$$

where r_i represents the red channel component ; g_i represents the green channel component ; b_i represents the blue channel component.

The color characterization of tobacco leaves under HSV channel and LAB channel is similar to the above formula [10].

4.2 Correlation analysis between image features and sensory evaluation features

The extracted image features contain certain elements that exhibit low correlation with the taste of tobacco leaves [11]. To accurately identify the features strongly associated with the sensory evaluation of tobacco leaves, the correlation between the image features and sensory evaluation data was examined.

Before the correlation analysis, in order to eliminate the influence of unit inconsistency between different characteristics of tobacco leaves, the maximum and minimum normalization method was used to standardize the data, and the data were mapped to the interval of [0,1], which is convenient for comparison under the same dimension [12]. The standardization formula is shown in (5):

$$X=\frac{x-\min(x)}{\max(x)-\min(x)} \tag{5}$$

Following standardization, correlation analysis was conducted between image features and sensory evaluation metrics (aroma quality, aroma quantity, offensive odor, irritation, aftertaste) for each grade of tobacco leaves. Subsequently, 10 features exhibiting a correlation coefficient of ≥0.25 with sensory evaluation metrics were identified as the definitive image features for tobacco leaves [13]. The correlation analysis outcomes for these 10 features were illustrated in Figure 7, denoted by serial numbers 3, 4, 6, 8, 9, 12, 13, 18, 22, and 23, respectively.

	3	4	6	8	9	12	13	18	22	23
1	-0.19	-0.19	-0.19	-0.26	-0.22	-0.21	-0.23	0.17	0.35	-0.18
2	0.22	0.22	0.053	0.12	0.19	0.18	0.18	0.25	0.21	0.053
3	0.24	0.24	0.28	0.21	0.23	0.28	0.27	0.22	-0.0078	0.34
4	0.29	0.29	0.28	0.26	0.28	0.27	0.26	0.47	0.27	0.31
5	0.16	0.19	0.17	0.16	0.18	0.16	0.17	0.23	0.13	0.24

Fig. 7. Correlation analysis between image features and sensory evaluation of tobacco leaves.

5 Feature fusion

In the study, we evaluated the visual and near-infrared spectral characteristics of tobacco leaves [14]. To enhance the assessment of tobacco leaf similarity, the features extracted from these two modes were combined to capture the unique qualities of each leaf. By utilizing objective data to measure leaf similarity, we aim to replace the conventional leaf grading formula that relies on sensory evaluation and expert judgment, offering a faster and more accurate method for finding similar tobacco leaves [15].

In the feature fusion phase, we compared four fusion methods: simple concatenation, multiplication, attention mechanism, and tensor fusion [16]. Simple concatenation involves combining multiple features into a larger vector based on a specific dimension. Multiplication fusion computes the combined feature representation by multiplying corresponding elements across different features. The attention mechanism fusion constructs an adaptive function to assign varying weights to features, emphasizing more relevant and significant information. This process can be summarized as:

$$Y = F_{sa}^{attention}\left[Sigmoid\left(F_{\max}, F_{avg}\right)\times W + b, X\right] \tag{6}$$

Tensor fusion involves merging multiple tensors to create a new tensor, characterized as the vector field of the triple Cartesian product, as illustrated in Formula (7) [17]:

$$\left\{\left(Z^{l}, Z^{v}, Z^{a}\right) \mid Z^{l} \in \begin{bmatrix} Z^{l} \\ 1 \end{bmatrix}, Z^{v} \in \begin{bmatrix} Z^{v} \\ 1 \end{bmatrix}, Z^{a} \in \begin{bmatrix} Z^{a} \\ 1 \end{bmatrix}\right\} \tag{7}$$

The added constant dimension is assigned a value of 1, allowing for the differentiation between single-modal and dual-modal data [18]. Each neural network coordinate (Z^{l}, Z^{v}, Z^{a}) can be visualized as a three-dimensional point within the triple Cartesian space, defined by near-infrared spectral features, image features, and subjective features represented in the dimensions $[z^{l}1]^{T}$, $[z^{v}1]^{T}$, and$[z^{a}1]^{T}$.

5.1 Evaluating indicator

As a unique agricultural commodity, tobacco leaves must undergo assessment based on their physical and chemical characteristics to ensure they align with the desired stability and flavor profile of the blend [19]. The evaluation process focuses on six key parameters - nicotine content, total sugar content, reducing sugar content, total nitrogen content, potassium content, and chlorine content, to determine the quality of tobacco leaves. Additionally, the sensory evaluation of tobacco leaves was conducted by seven experienced smoking experts at the redrying factory [20]. This study compared the differences between the primary tobacco material and potential replacement tobacco leaves from both a physical and chemical perspective. The methodology for assessing the similarity of tobacco leaves is further elucidated through the mathematical formulations presented in equations (8), (9), and (10).

$$PC = \left|(a-a')\right| + \left|(b-b')\right| + \left|(c-c')\right| + \left|(d-d')\right| + \left|(e-e')\right| + \left|(f-f')\right| \tag{8}$$

$$SE = \left|m-m'\right| \tag{9}$$

$$DV = 0.5PC + 0.5SE \tag{10}$$

In this context, PC signifies the discrepancy in physical and chemical properties, with a, b, c, d, e, f representing the standard chemical content values of nicotine, total sugar, reducing sugar, total nitrogen, potassium, and chlorine of the specific primary tobacco material as determined by the prescribed continuous flow method outlined in relevant standards like YC/T 159-2002. Meanwhile, a', b', c', d', e', and f' denote the standard chemical content values of nicotine, total sugar, reducing sugar, total nitrogen, potassium, and chlorine for the potential replacement tobacco leaves. SE, on the other hand, indicates the distinction in sensory properties, with m representing the sensory score of the primary tobacco material and m ' referring to the sensory score of the substitute tobacco. DV ultimately denotes the overall comprehensive difference value.

5.2 results comparison

We used Euclidean distance to measure the similarity between sample points. This distance metric calculates the straight-line distance between two points in the Cartesian coordinate system, making it a common measure of proximity. A smaller Euclidean distance indicates a closer spatial relationship between two vectors, implying greater similarity. The Euclidean distance d between two n-dimensional points $P = (p_1, p_2, \cdots, p_n)$ and $Q = (q_1, q_2, \cdots, q_n)$ can be calculated using Formula (11):

$$d(P,Q) = \sqrt{(p_1-q_1)^2 + (p_2-q_2)^2 + \cdots + (p_n-q_n)^2} \tag{11}$$

Assuming the target tobacco material is A-C2F-CO2, similar leaves are identified among 24 others. Using Euclidean distance in low latitude, we employed four fusion methods—simple splicing, multiplication, attention mechanism, and tensor fusion—to identify replacements for A-C2F-CO2 leaves. The top three replacements, determined by Euclidean distance, are compared in Table 2.

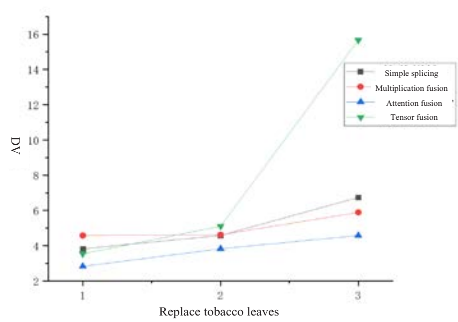


Fig. 8. Comparison of DV values of different fusion methods.

Table 2. Comparison of replacement tobacco leaves obtained by different fusion methods.

Method	Unblended cigarette	Chemical composition						Sensory score	DV
		Nicotine	Reducing sugar	Total sugar	Total nitrogen	Potassium	Chlorine		
	Target	2.69	23.35	35.09	2.184	1.195	1.96	83.33	
Simple splicing	Replace 1	2.528	21.954	32.21	2.002	1.294	1.895	80.27	3.824
	Replace 2	2.759	20.51	29.85	2.275	1.904	1.668	84.99	4.582
	Replace 3	1.621	26.6	37.58	1.761	0.396	2.023	78.49	6.738
Multiplication fusion	Replace 1	2.759	20.51	29.85	2.275	1.904	1.668	84.99	4.582
	Replace 2	2.522	22.14	30.14	2.106	0.196	1.796	81.66	4.62
	Replace 3	2.196	24.36	29.58	2.032	0.226	1.854	77.77	5.891
Attention fusion	Replace 1	2.808	21	31.291	2.443	2.187	2.293	84.55	**2.834**
	Replace 2	2.528	21.954	32.21	2.002	1.294	1.895	80.27	3.824
	Replace 3	2.759	20.51	29.85	2.275	1.904	1.668	84.99	4.582
tensor fusion	Replace 1	1.722	24.567	32.356	1.994	0.498	2.147	79.44	3.54
	Replace 2	3.782	22.7	27.98	2.898	0.928	1.858	79.44	5.107
	Replace 3	2.157	15.06	17.34	2.637	0.507	3.007	77.77	15.661

From Table 2 and Figure 8, we observe the difference values for replacement tobacco leaves obtained through various fusion methods. The smallest difference value, 2.8335, is achieved by the attention mechanism fusion method, indicating its superiority in preserving leaf features. This method also yields the lowest comprehensive difference values for replacement leaves, prioritizing their selection. Hence, we opted for the attention mechanism fusion method for multi-modal feature fusion.

Using this approach, we calculated the similarity values between the target tobacco leaf A-C2F-CO2 and the remaining 24 replacement leaves. Denoted as M, these values were computed using Euclidean distance, where smaller values denote greater similarity. Table 3 presents the multi-modal similarity results.

Table 3. Multimodal similarity value of tobacco leaves.

Numbering	Unblended cigarette	Similarity M	Numbering	Unblended cigarette	Similarity M
1	A-C2F-C02	0.0000	4	A-C3L-C03	0.8605
2	A-C2F-C03	0.2001	17	D-C3F-C03	0.8675
19	E-C2F-C03	0.5472	18	D-C4F-C03	0.8859
7	B-C2F-C03	0.6390	22	F-B1F-B01	0.9067
8	B-C3F-C02	0.6811	9	B-C3F-C03	0.9085
3	A-C3F-C03	0.6817	16	D-C3F-C02	0.9111
23	F-C2F-C03	0.6932	12	C-C3F(1)-C03	0.9707
11	C-C2F(1)-C03	0.7333	24	F-C3F-C03	0.9868
10	C-B1F(1)-B01	0.7545	5	A-C4F-C03	1.2242
15	D-C2F-C03	0.7658	13	C-C4F(1)-C03	1.5456
20	E-C3F-C03	0.7889	21	E-C4F-C03	1.6375
6	B-B1F-B01	0.8063	25	F-C4F-C03	1.6741
14	D-C2F-C02	0.8462			

6 Conclusion

This study proposes a method for calculating tobacco leaf similarity based on multi-modal fusion. It employs a 1D-CNN network to extract near-infrared spectral features and utilizes relevant functions in OpenCV for extracting image features of tobacco leaves. Correlation analyses are then conducted between these features and sensory scores of tobacco leaves.

Features highly correlated with tobacco taste are chosen as the near-infrared spectral and image features. Four different fusion methods are compared, ultimately selecting the attention mechanism feature fusion method to fuse the near-infrared spectral and image features. Tobacco leaf similarity is computed using the Euclidean distance, obtaining similarity values between the target tobacco leaf and potential replacements. This facilitates the identification of the most similar tobacco leaf. The method enables objective judgment and intelligent recommendation for substituting individual tobacco leaves when a particular ingredient is insufficient in cigarette leaf blending, thereby maintaining the inherent quality of the blend.

References

1. X. Shen, Stat. Appl., **05**, 28-38 (2016)
2. F. Fan, S. Wang, C. Wang, Sensors, **19**,23(2019)
3. Y. Li, N. Wang, J. Shi , Pattern Recognit. J. Pattern Recognit. Soc., **80**, 109-117(2018)
4. D. Wang, F. Tian, Z. Zhu, Sensors ,**20**, 874(2020)
5. S. Zhu, L. Zhou, P. Gao, Molecules, **24**, 3268(2019)
6. A. Zadeh, M. Chen, S. Poria, arXiv, (2017)
7. Y. Zhang, Q. Cong, Y. Xie, Spectrochim. Acta. A. Mol. Biomol. Spectrosc., **71**, 1408-1413(2009)
8. X. Shen, Stat. Appl., **05**, 28-38(2016)
9. D. Giokas, N. Thanasoulias, J. Hazard. Mater., **185**, 86-92(2011)
10. 10. J. Hou, Y. Ma, C. Lu, Spectrochim. Acta. A. Mol. Biomol. Spectrosc., **251**, 119364(2021)
11. Y. Bi, S. Li, Spectrochim. Acta. A. Mol. Biomol. Spectrosc., **215**, 398-404(2019)
12. N. Bowen, He. Minyi, B. Cao, Journal of Medical Internet Research., **23**, e25700(2021)
13. A. Bhuyan, K. Sanguri, H. Sharma, *2021 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, 482-487(2021)
14. W. Wu, Z. Lv, L. Zhao, Sci. Manag. Res., **32**, 74-77(2014)
15. J. Xin, Z. Niu, J. Huang, IEEE Trans. Pattern Anal. Mach. Intell., **40**, 987-1001(2018)
16. C. Zhang, Q. Hu, H. Fu, IEEE, **461**(201)
17. G. Peng, Z. Jiang, H. You, (2018)
18. S. Ren, K. He, R. Girshick, IEEE Trans. Pattern Anal. Mach. Intell., **39**,1137-1149(2017)
19. K. Cho, V. Merrienboer, C. Gulcehre, Comput. Sci., (2014)
20. N. Xu, W. Mao, P. Wei, IEEE Intell. Syst., **36**,3-12(2021)