



## A novel framework designed by MMS-UNet and DK-SAC for the smart systems of the traditional Chinese medicine manufacturing

Zhongxin Chen, Yu Chen, Yongwei Tang, Zhijian Wang, Yanfeng Li, Lei Dong & Weibo Ren

**To cite this article:** Zhongxin Chen, Yu Chen, Yongwei Tang, Zhijian Wang, Yanfeng Li, Lei Dong & Weibo Ren (05 Mar 2025): A novel framework designed by MMS-UNet and DK-SAC for the smart systems of the traditional Chinese medicine manufacturing, *Journal of Engineering Design*, DOI: [10.1080/09544828.2025.2473983](https://doi.org/10.1080/09544828.2025.2473983)

**To link to this article:** <https://doi.org/10.1080/09544828.2025.2473983>



Published online: 05 Mar 2025.



Submit your article to this journal 



View related articles 



CrossMark

View Crossmark data 



# A novel framework designed by MMS-UNet and DK-SAC for the smart systems of the traditional Chinese medicine manufacturing

Zhongxin Chen <sup>b</sup><sup>a</sup>, Yu Chen<sup>b</sup>, Yongwei Tang<sup>c</sup>, Zhijian Wang<sup>a</sup>, Yanfeng Li<sup>a</sup>, Lei Dong<sup>a</sup> and Weibo Ren<sup>a</sup>

<sup>a</sup>The School of Mechanical Engineering, North University of China, Taiyuan, People's Republic of China; <sup>b</sup>The Information Technology Center, Qingdao Port International Logistics Co., Ltd, Qingdao, People's Republic of China; <sup>c</sup>Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Qilu University of Technology (Shandong Academy of Sciences), Jinan, People's Republic of China

## ABSTRACT

The traditional Chinese medicine (TCM) industry is constantly exploring the application of service-oriented smart manufacturing (SoSM) systems to meet the requirements of Pharmaceutical 4.0. However, traditional manufacturing systems cannot achieve online detection and self-adaptive control (SAC) of critical quality attributes (CQAs), which plays an important role in the interaction between the physical and service worlds in the SoSM. Therefore, we propose the multi-modal similarity UNet (MMS-UNet) to collect features from multispectra data fusion, particularly focusing on the online CQAs data acquisition task. To achieve more accurate detection, we encoded spectral data as images using segmented Gramian angular fields (SGAF). The experimental results demonstrate the superiority of the proposed method: it achieves higher detection accuracy and stability, as indicated by the mean absolute error (MAE). Additionally, based on the rapid CQAs data, we designed a data-knowledge hybrid driven SAC (DK-SAC) framework to improve quality and batch consistency in TCM. By addressing the specific challenges of Xiao'er Xiaoji Zhike Oral Liquid (XXZOL), the research aims to promote process optimisation and intelligent control in the pharmaceutical industry. The proposed framework is expected to fill the data exchange gap in SoSM systems and reduce regulatory risks in the pharmaceutical industry.

## ARTICLE HISTORY

Received 3 November 2024

Accepted 26 February 2025

## KEYWORDS

Data-knowledge hybrid driven; data fusion; self-adaptive control; traditional Chinese medicine

## 1. Introduction

The traditional Chinese medicine (TCM) production industry, directly related to human life safety, has always been a regulatory focus for government agencies (Lei et al. 2024). Although strict quality supervision ensures the safety of drugs, it also suppresses the industry's innovation capabilities. This presents a challenge to implements Pharmaceutical

---

**CONTACT** Weibo Ren 20230093@nuc.edu.cn The School of Mechanical Engineering, North University of China, Taiyuan, Shanxi, People's Republic of China

Industry 4.0 (Keung et al. 2024; Siiskonen et al. 2023). Fortunately, the rapid development of service-oriented smart manufacturing (SoSM) technology has greatly enhanced the intelligence of the pharmaceutical industry (Zhou et al. 2023). First, the typical combination of artificial intelligence and robotics enables the system to operate smoothly without human involvement, thereby reducing the risk of drug contamination. Second, the Internet of Things (IoT) and big data technologies can integrate multiple data sources, achieve interconnectivity of internal and external data, and enhance production efficiency. These aspects validate the inevitable trend of implementing Pharmaceutical Industry 4.0, which is fully supported by the disruptive applications brought by digital integration technology in the IoT fields of chemical pharmaceuticals and biopharmaceuticals, such as real-time, on-demand, and self-adaptive production.

The foundation of SoSM lies in collecting various data during the production process through modern sensing technologies, enabling real-time monitoring and data collection from production equipment. However, the critical quality attributes (CQAs), which are chemical substances that directly affect the efficacy of drugs, are difficult to obtain in real-time due to the numerous raw materials and the complex chemical environments in TCM manufacturing (Ma et al. 2023). Additionally, the lack of scientifically reliable real-time analysis, detection, and prediction methods hinders the optimisation of critical process parameters (CPPs) control strategies (Gao et al. 2023; Yenduri et al. 2022). In this regard, infrared spectroscopy technology can perform rapid quantitative and qualitative analysis of chemical components in complex mixtures through non-invasive methods, thereby facilitating the online detection of CQAs in the TCM process (K. Wang et al. 2025).

Different types of infrared spectra have significant differences in wavelength ranges, fields of action, and application equipment. For example, near-infrared spectroscopy (NIRS) records the absorption peaks of hydrogen-containing functional groups at multiple combined frequencies, while mid-infrared spectroscopy (MIRS) mainly reflects the fundamental frequency vibration characteristics of certain functional groups in organic compounds. However, a single infrared spectroscopy analysis technique cannot provide comprehensive information about samples containing multiple organic compounds. The use of multi-source data fusion technology can compensate for the shortcomings of a single detection method and provide more detailed information (Q. Zhang et al. 2024). For example, the data-level fusion method was established for NIRS-MIRS based on dynamic orthogonal projection (Liu et al. 2024). The data fusion schemes that combine NIRS and MIRS data were constructed to conduct a quantitative analysis of the illegal addition of atenolol in Panax notoginseng (Du et al. 2024). However, multi-source data fusion increases the complexity of data processing and computational costs (Dong et al. 2025). UNet, one of the successful deep learning architectures, has recently shown promising results in this area because of the unique contraction and expansion paths to learn the multi-source features in an end-to-end fashion (Hu et al. 2019).

On its own, infrared spectroscopy technology can provide real-time information for online quality detection in the TCM process and help researchers gain a comprehensive understanding of the changing trends of CQAs. On the other hand, with the help of CQAs data, production time can be adaptively set based on quality data, altering the original production process from a fixed time standard and achieving a more efficient and smarter TCM manufacturing. However, obtaining CQAs from online data and guiding the optimisation

of TCM processes remain a major challenge. In summary, the main research questions of this article are as follows:

- RQ1: How can we build a data fusion analysis system based on complex online production conditions, especially by combining NIRS with MIRS, to accurately and efficiently obtain CQAs data, which is the key basic data of SoSM in the TCM process?
- RQ2: How can the CQAs data guide the processes of TCM and design a control framework to simultaneously meet the stringent regulatory requirements and the self-adaptive control (SAC) requirements of smart manufacturing?

In response to the above questions, this study mainly contributes to existing work in two ways:

- (1) We propose a multi-modal similarity UNet (MMS-UNet) architecture for NIRS and MIRS that uses the spatial distribution similarity between different types of spectra to fuse multi-modal features. This architecture provides a new paradigm for obtaining CQAs data through online NIRS in TCM processes.
- (2) A data-knowledge hybrid driven SAC(DK-SAC) framework has been constructed based on the obtained online CQAs data. In the framework, historical data and empirical knowledge are defined as supplementary features to guide different control decisions, which further facilitates the representation of quality knowledge in a control design context, demonstrating effectiveness in risk management and efficient production.

## 2. Related work

### 2.1. Research on online detection in TCM manufacturing

In TCM manufacturing, CQAs testing methods are mostly based on sampling and end-of-pipe testing, lacking scientific and reliable real-time analysis, monitoring, and prediction methods. NIRS technology has been widely used in chemical analysis (Gan and Luo 2023; Robert and Gosselin 2022), food science (An et al. 2023; Lukacs et al. 2018), neurocognitive mechanisms (F. Wang et al. 2024), and pharmaceutical fields (Deidda et al. 2019). NIRS, due to its fast, non-destructive, and efficient characteristics, is particularly effective as a detection tool for the online detection of CQAs in TCM (Lei and Sun 2020; Robert and Gosselin 2022). NIRS was applied to different batches of the Flos Lonicerae Japonicae extraction process, and a new calibration model transfer strategy was proposed to maintain the predictive abilities for a new source (A. Wang et al. 2019). The CQAs identification strategy and rapid quantitative evaluation of Flos trollii were constructed based on NIRS and chemometric methods (Mao et al. 2024). The electronic eye and NIRS were used to explore the changes in CQAs for Epimedium, and the method combining convolutional neural networks (CNN) with partial least squares (PLS) established the qualitative discrimination and prediction model (J.-B. Zhang et al. 2024). Researchers used MIRS to indirectly analyse CQAs of Danshen granules (Guo et al. 2016) and the geographical origin of Gentiana scabra in southwestern China (Liu et al. 2020). It is worth noting that CQAs information

from complex medicinal solutions can be obtained by using infrared spectroscopy technology. However, the process environment of TCM is complex, and the contribution of a single infrared spectroscopy technology to understanding the comprehensive information on CQAs is limited.

Fortunately, the method of multi-source information fusion can not only utilise the characteristics of different spectral technologies but also provide more comprehensive information for TCM process. In recent years, multispectral data fusion technology has emerged, integrating different types of spectral data to achieve information synergy and complementarity, making the prediction results of constructed qualitative or quantitative analysis models more reliable and accurate (Hao et al. 2024). For example, MIRS and NIRS have been combined with data fusion strategies to address traceability (Lan et al. 2020) and adulteration issues (Du et al. 2024) in TCM. In addition, these two technologies combined with data fusion strategies have been used to quantitatively determine the content of CQAs in *Lonicera japonica* and *Artemisia annua* (Tao et al. 2019).

Most existing research on TCM focuses on one or a few chemical components. Because TCM is extracted from a mixture of various natural products, it is difficult to accurately evaluate its quality by predicting the levels of multiple chemical components. Fortunately, most chemical components can be detected by high performance liquid chromatography (HPLC). Therefore, this study adopts a data fusion strategy to convert NIRS and MIRS data into more intuitive spectral data to obtain real-time CQAs data during the manufacturing process.

## 2.2. Research on optimisation of control systems in TCM manufacturing

The application of machine learning (ML) algorithms to optimise control technology in TCM manufacturing and achieve intelligent control is a growing trend. In industrial manufacturing, it has been fully demonstrated that the application of ML in process optimisation and control is effective (Jiao et al. 2022; Nakata et al. 2017). Compared with other industrial processes, the TCM manufacturing process pays more attention to online control based on CQAs. Researchers are exploring more effective ML methods through experimentation. Regarding the nonlinear time-varying dynamic system in the fermentation process of pharmaceutical production, researchers collected 40 batches of data, and applied multiple ML algorithms to optimise the production process, identify key influencing factors and apply the results to production line optimisation (Sun et al. 2019). Model predictive control proportional integral derivative (MPC-PID) was developed and implemented to achieve stable CQAs control for continuous tablet manufacturing (Singh et al. 2014). The above research is mainly based on control methods using empirical or expert knowledge, which is reflected in the use of fixed process parameters to control the processes, which is currently the common control mode in TCM manufacturing (Zhu, Ouyang, and Kong 2024). The biggest advantage of this approach is that it can meet strict regulatory requirements, but it cannot meet the requirements of intelligent manufacturing and seriously affects batch consistency in the process (Q. Wang et al. 2023).

In the design of smart manufacturing systems, researchers use self-adaptive control (SAC) methods to adjust various industrial parameters in real time, improving the autonomy, adaptability, and scalability of the entire process (Brooks and Roy 2021). This effective solution was proposed, requiring the setting of the model's state, control input signals, and

real-time production process data as feedback signals (Kang et al. 2024). PLS was used to establish the relationship between material properties, process parameters, and CQAs, and optimise process parameters by controlling the CQAs content (Garcia-Munoz and Mercado 2013). The multiple control strategy based on the Bayesian paradigm was established and applied to a quality-adaptive control system in drug infusion manufacturing systems (Radcliffe and Reklaitis 2021). Some scholars designed the SAC optimisation strategy for CPPs by analysing 90 influencing factors and the dissolution rate of tablets containing two active ingredients using PLS (Prikernik and Srcic 2021). Most of the above studies are based on simple or single CQA-controlled pharmaceutical processes, and their contribution to SAC in the complex process of TCM manufacturing is limited.

In summary, the advantages of SAC systems based on ML algorithms cannot be ignored and are sufficient to change the current paradigm of TCM manufacturing systems design. This method can break through the limitations of fixed processes and truly achieve intelligent, real-time CQAs control, which is key to promoting SoSM in TCM manufacturing.

### 3. The designed architecture and key methods

This section explains how to design an SAC architecture for TCM manufacturing using the online multifunctional NIRS acquisition system, as shown in Figure 1. This study designed the DK-SAC framework based on the proposed MMS-UNet method and promoted intelligent control in TCM manufacturing.

#### 3.1. Developed online CQAs data acquisition algorithm

To assist relevant researchers in better utilising analytical chemistry technology and artificial intelligence for the online detection of CQAs, this study constructed an intelligent algorithm framework based on multi-source data to realise its engineering application value.

First, to fully utilise the advantages of deep learning algorithms in feature extraction from two-dimensional images, while preserving the continuity of one-dimensional original spectroscopy data and the correlation information between wavelength and absorption values, this paper adopts the GAF method to convert one-dimensional spectra into two-dimensional images.

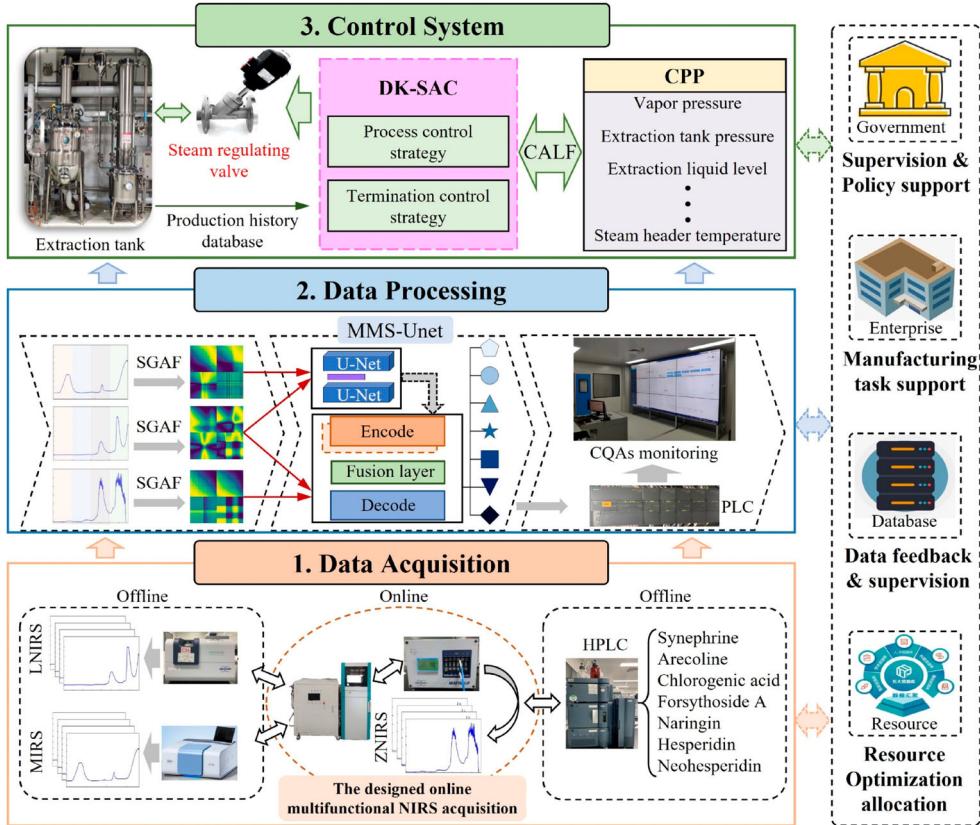
The spectroscopy data,  $X = \{x_1, x_2, \dots, x_n\}$ , is scaled in  $[-1, 1]$  by Equation (1).

$$\tilde{x}_{-1}^i = \frac{(x_i - \max(X)) + (x_i - \min(X))}{\max(X) - \min(X)} \quad (1)$$

The polar coordinate transformation and inverse cosine function are used to encode the absorbance value as an angle value, and the wavenumber is encoded as a radius  $r$ . The spectroscopy data  $X$  in the Cartesian coordinate system is converted into  $\tilde{X}$  in polar coordinates using the Equation (2):

$$\begin{cases} \theta = \arccos(\tilde{x}_i), & -1 \leq \tilde{x}_i \leq 1, \tilde{x}_i \in \tilde{X} \\ r = \frac{t_i}{N}, & t_i \in N \end{cases} \quad (2)$$

where,  $N$  is the average number of parts dividing the scale interval  $[-1, 1]$ .



**Figure 1.** The architecture of CQAs data acquisition method and DK-SAC framework.

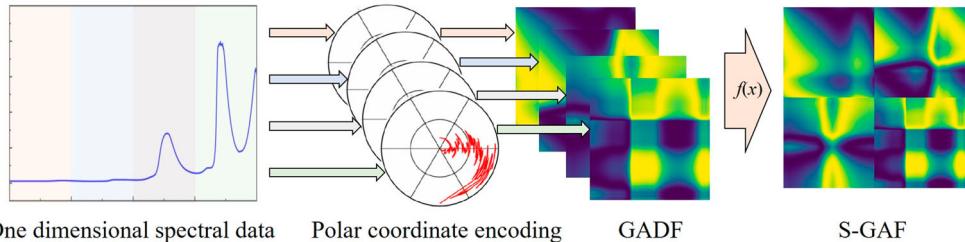
We generate a gramian angular difference field (GADF) based on a sine function using Equation (3).

$$\text{GADF} = G(i,j) = \begin{bmatrix} \sin(\theta_1 + \theta_1) & \cdots & \sin(\theta_1 + \theta_j) \\ \vdots & \ddots & \vdots \\ \sin(\theta_i + \theta_1) & \cdots & \sin(\theta_i + \theta_j) \end{bmatrix} = \sqrt{I - \tilde{X}^2} \tilde{X} - \tilde{X}^T \sqrt{I - \tilde{X}^2} \quad (3)$$

where,  $I$  is the unit row vector  $[1, 1, \dots, 1]$ .

The absorbance at different wavenumbers is converted into a sine of the angle difference, associating all wavenumber information with absorbance information and preserving both wavenumber and absorbance information in the original spectroscopy. Based on the monotonicity of the  $\sin(\theta)$ , using GAF conversion has two main advantages: first, it provides a method to maintain wavelength dependence in spectroscopy data (as shown in Equation (3), wavelength information starts from the lower left corner of the image and moves to the upper right corner, with each pixel containing wavelength information); Another point regarding spectroscopy information is that the value of each pixel represents whether the information at that location is additive or differential within the wavelength range.

For one-dimensional spectra, there are significant differences in absorbance across different wavenumber regions, and the correlation between variables in these different



**Figure 2.** Workflow of SGAF.

wavenumber regions also needs to be considered during the spectral conversion process. Based on the extreme gradient boosting (XGBoost) and GAF algorithm, this article calculates the feature weight of different wavenumber regions. This weight is the average gain brought by the feature as a splitting node in all regression trees. Therefore, the characteristic values for different wavenumber regions can be expressed as:

$$W_l = \frac{1}{2} \frac{\sum_{t=1}^T \sum_{i=1}^{N(t)} I(\beta(t, i) = k) \left( \frac{G_{\gamma(t, i, L)}^2}{H_{\gamma(t, i, L)} + \lambda} + \frac{G_{\gamma(t, i, R)}^2}{H_{\gamma(t, i, R)} + \lambda} - \frac{G_{\gamma(t, i)}^2}{H_{\gamma(t, i)} + \lambda} \right)}{\sum_{t=1}^T \sum_{i=1}^{N(t)} I(\beta(t, i) = k)}, \quad l = 1, 2, 3, 4 \quad (4)$$

Where,  $k$  represents a certain node,  $T$  represents the number of all trees,  $N(t)$  represents the number of non leaf nodes in the  $t$ th tree,  $\beta(t, i)$  represents the partition feature of the  $i$ th non leaf node in the  $t$ th tree, and  $\beta(g)$  is the indicator function.  $I$  represents different wavenumber regions.

Then, this article proposes a segmented GAF (SGAF) algorithm for converting spectral data, as shown in Figure 2. The algorithm divides each spectral dataset into four segments, calculates the weight of each wavenumber region according to Equation (4) after GAF, and merges them into one graph. The image is assigned colour information using the colormap method and the 'Viridis' is selected. The pixel size of the two-dimensional spectral image generated by the SGAF transformation is  $640 \times 640$ .

Second, using the SGAF algorithm to obtain 2D images from 1D spectral data, we further need to establish a prediction model for CQAs through two stages.

In the first stage, the fusion of offline NIRS (LNIRS) and MIRS was involved, and two different UNet models were independently trained using LNIRS and MIRS, respectively. This article utilises MIRS information to assist LNIRS-UNet in learning NIRS features and accelerates the feature extraction process of the network. Specifically, the encoding part of LNIRS-UNet is the same as MIRS-UNet, with the only difference being the feature extraction stage of the encoder, where the authors use squeeze and excitation (SE block) (T. Zhang and Zhang 2022) to obtain the attention map of MIRS information, and then multiply it element by element with the LNIRS feature map. This process can inject significant information captured by MIRS into LNIRS features to assist the LNIR-UNet network in feature extraction. The original UNet architecture is in the first stage.

The second stage involves data fusion and reconstruction of LNIRS and online NIRS (ZNIRS). This article adopts the method of feature layer fusion to form feature vectors and uses specific network architecture pattern recognition methods to extract deep semantic

features as a basis for further prediction of CQAs. This article uses the encoder of LNIRS-UNet to extract the shallow and deep semantic features of LNIRS. Then, the first four layers of the encoder of LNIRS-UNet are trained on ZNIRS, and LNIRS-UNet and the first four layers of ZNIRS-UNet are fused to generate four sets of feature maps for decoder training. For the fusion strategy, the feature maps of LNIRS are down-sampled to match the size of ZNIRS, which amplifies important features and suppresses irrelevant features.

The authors introduced a new loss function (Chen et al. 2023) to achieve the decomposition and reconstruction, using the following formula:

$$\text{PDS}(X, Y) = \frac{\sum_{m=1}^n (X_{i,m} - \bar{X})(Y_{j,m} - \bar{Y})}{\sqrt{\sum_{m=1}^n (X_{i,m} - \bar{X})[\sum_{m=1}^n (Y_{j,m} - \bar{Y})]}} \quad (5)$$

where,  $\bar{X}$  represents the mean of image  $X$ ,  $\bar{Y}$  represents the standard deviation of image  $Y$ . Therefore, the calculation formula of the structural loss,  $\text{Loss}_{\text{PDS}}$ , is as follows:

$$\text{Loss}_{\text{PDS}} = \text{PDS}(Z, X) + \text{PDS}(Z, Y) \quad (6)$$

$$\text{Loss}_{\text{det}} = \nabla Z - \max(\nabla X, \nabla Y)^2 \quad (7)$$

$$\begin{aligned} \text{Loss} = & \text{Loss}_{\text{mse}}((\text{LNIRS} + \text{ZNIRS})/2, F) + \alpha \text{Loss}_{\text{PDS}}(\text{LNIRS}, \text{ZNIRS}, F) \\ & + \beta \text{Loss}_{\text{det}}(\text{LNIRS}, \text{ZNIRS}, F) \end{aligned} \quad (8)$$

where,  $\nabla$  is a sobel operator, and  $F$  represents the fused image.

Then, this article obtains effective feature maps through the above fusion strategy, and further predicts CQAs using PLS algorithm. The overall workflow of MMS-UNet is shown in Figure 3.

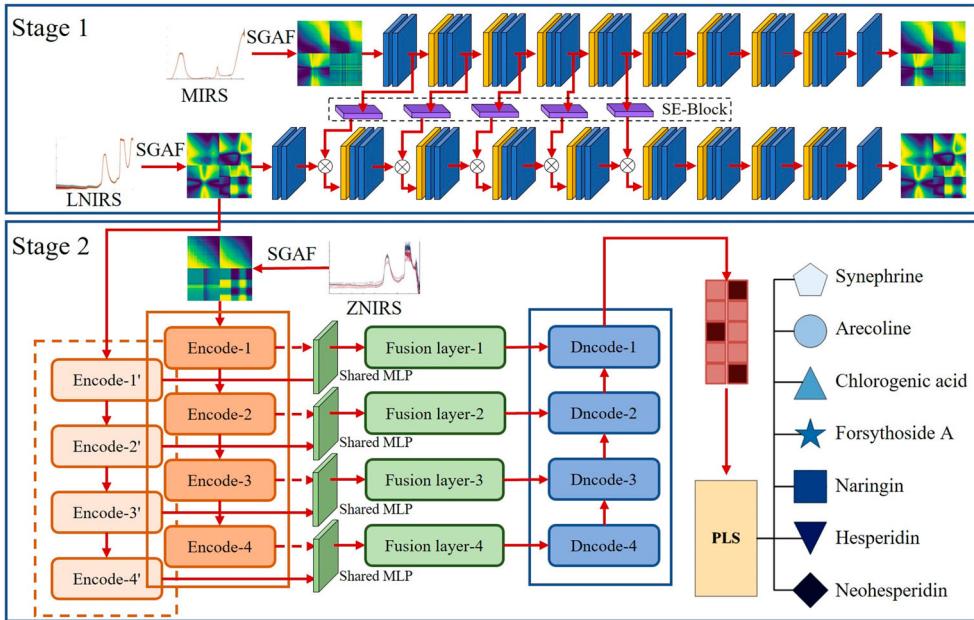
### 3.2. The DK-SAC system for TCM manufacturing

CPPs are a direct factor affecting the content of CQAs in the TCM process. In the previous section, we achieved real-time prediction of CQAs data based on the proposed MMS-UNet algorithms, breaking through the barriers of key data interaction between the physical world and the information world in SoSM. Therefore, this section takes the control of steam regulating valve opening as the research object and proposes a DK-SAC algorithm based on data knowledge hybrid drive, which breaks the original fixed process parameter method and meets the strict quality requirements of TCM manufacturing.

The steam regulating valve controls the pressure, temperature, and extraction time of the multifunctional extraction tank and is the core controller of the system. The overall architecture of the control system is shown in Figure 1, which controls the opening of the steam regulating valve based on the extraction process and extraction endpoint.

This article is based on the CALF algorithm (Chen et al. 2022), and establishes a correlation model between CPPs and CQAs in the extraction process. The production process model can be represented as:

$$\begin{cases} C_i(t) = f^i(X(t)) = \frac{1}{K} \sum_{k=1}^K f_{E_k}^i(X(t)) = \frac{1}{K} \sum_{k=1}^K \left[ f_{k_0}^i(X(t)) + \sum_{e=1}^E \sum_{s=1}^S c_{es}^i I(X(t) \in R_{es}^i) \right] \\ X(t) = [T(t), W(t), J(t), T_L(t), T_Y(t), T_Z(t), Z(t), G(t)] \end{cases} \quad (9)$$



**Figure 3.** Workflow of MMS-UNet.

Where,  $C_i(t)$  represents CQAs content,  $i = 1, 2, \dots, 7$ ;  $T(t)$  represents extraction time,  $W(t)$  represents liquid level,  $J(t)$  represents jacket steam pressure,  $T_L(t)$  represents circulating water temperature,  $T_Y(t)$  represents liquid temperature,  $T_Z(t)$  represents steam header temperature,  $Z(t)$  represents steam header pressure,  $G(t)$  represents extraction tank pressure, and  $K$  represents the number of base learners

The opening of the steam regulating valve is controlled by the difference between the CQAs content detected online and the CQAs content output from the production process model. The specific algorithm is as follows:

$$\Delta c_i = C_i(t_1) - c_i(t_2) \quad (10)$$

In the formula,  $t_1$  represents the real-time acquisition time of process parameters, with an interval of 1 s;  $t_2$  is the fast detection time for online CQAs, with an interval of 1 min;  $c_i(t)$  is the online CQAs content. The opening of the steam regulating valve is set to:

$$J'(t) = a + \sum_{i=1}^7 \varpi_i \Delta c_i \quad (11)$$

In the formula,  $a$  represents the initial setting value of the steam regulating valve,  $\varpi_i$  represents the proportional parameter, and is determined by the influence weights of different CQAs. During the extraction process, the opening of the steam regulating valve can be self-adaptive control based on data driven according to Equation (11).

To improve the quality consistency between batches in the process and meet the regulatory requirements, we propose a knowledge driven production endpoint control strategy based on historical data and expert knowledge. The parameters for setting the endpoint control strategy of the extraction process are as follows:

- (1) Maximum extraction time  $t_{\max}$  is set according to the Pharmacopoeia of the People's Republic of China (2020 edition).
- (2) Minimum extraction time  $t_{\min}$  is set to the time when the synephrine content detected online reaches its minimum value, which is specified in the Pharmacopoeia of the People's Republic of China (2020 edition).

In addition, the median values of the content intervals for the other six CQAs, except for Synephrine, are set as follows:

$$\bar{c}_i = \frac{(c_{i\max} + c_{i\min})}{2}, \quad i = 2, 3, \dots, 7 \quad (12)$$

where,  $c_{i\min}$ ,  $c_{i\max}$  represent the upper and lower limits of CQAs, then:

$$g_i(t) = [C_i(t) - \bar{c}_i]^2 \quad (13)$$

$$\tilde{g}(t) = \sum_{i=2}^7 g_i(t) \quad (14)$$

$$h(t) = \tilde{g}'(t) \quad (15)$$

According to the above formula, obtain a knowledge driven extraction termination control strategy.

In summary, based on the extraction process control strategy and extraction termination control strategy, the DK-SAC process of the steam regulating valve in process is as follows:

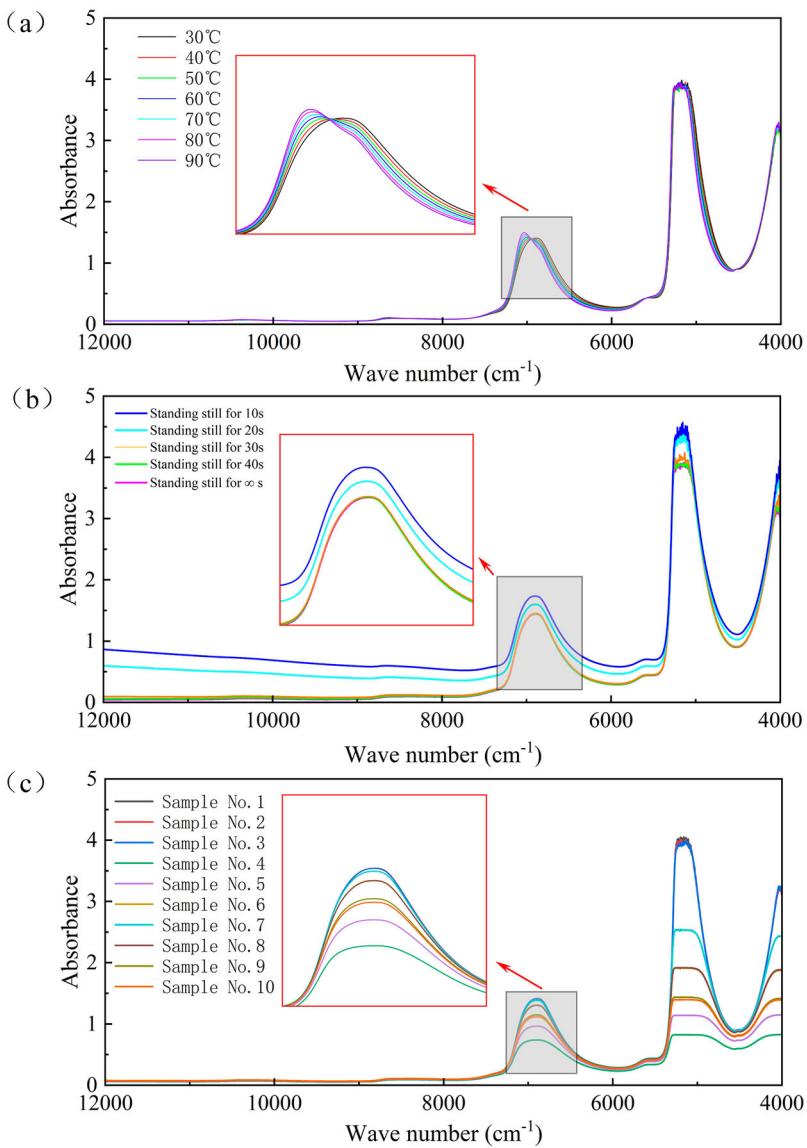
- (1) Set the CQAs content range index according to the extraction process requirements.
- (2) By measuring various CPPs data through sensors, using the CALF algorithm to output the CQAs content  $C_i(t_1)$ . The online CQAs rapid detection data was conculated by the MMS-UNet algorithm. The opening of the steam regulating valve was set according to equitation (11).
- (3) Use the extraction termination strategy to determine whether to set the steam regulating valve  $J'(t) = 0$ .

## 4. Methodology and case study

This section introduces a research design project that demonstrates the application of a data acquisition workflow in SoSM based on multi-source data fusion technology. In the field of TCM manufacturing, this case is a combination of computer science and spectroscopy analysis. The case should have both scale and represent typical data characteristics in the field. Therefore, we selected the Xiao'er Xiaoji Zhike Oral Liquid (XXZOL) from 95 major varieties of TCM as the subject, with a particular focus on the extraction process, which is the most common primary process in TCM manufacturing.

### 4.1. Design of an online multifunctional NIRS acquisition system

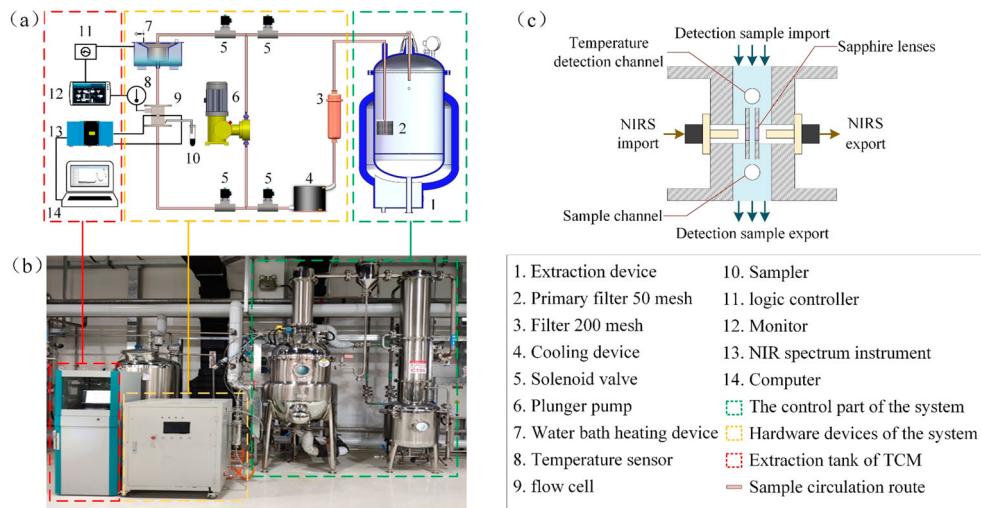
The online detection technology using NIRS as a detection tool is an effective way to achieve rapid online data collection of CQAs. The prerequisite for achieving stability and



**Figure 4.** The interference factors in NIRS: (a) Temperature; (b) Impurity content; (c) Bubbles.

effectiveness in online testing of manufacturing processes is to obtain high-quality samples. However, NIRS is sensitive to changes in the external environment, and the detection data are prone to interference. Therefore, we first need to study the impact of different interference factors on the quality of NIRS.

This article studied the effects of temperature, impurity content, and bubbles on NIRS, as shown in Figure 4. It is worth pointing out that this article describes the influence of impurity content on NIRS using the supernatant after stirring and allowing it to stand for different periods. It can be seen from Figure 4 that the influence of impurity content



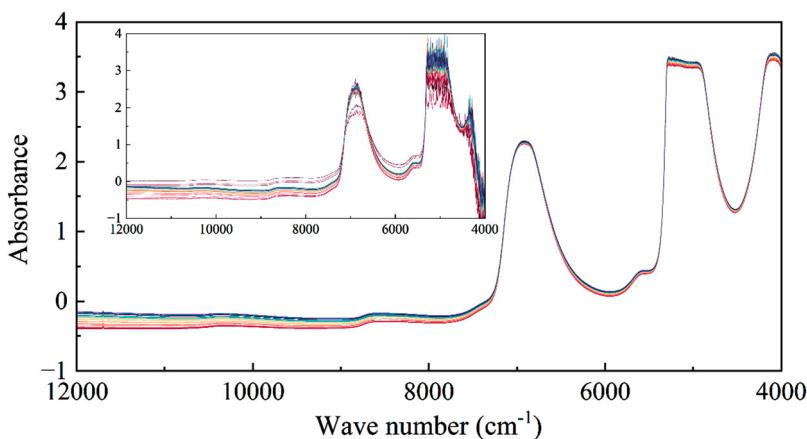
**Figure 5.** The designed online multifunctional NIRS acquisition system: (a) The designed scheme; (b) The application of the scheme; (c) The schematic diagram of flow cell.

on NIRS is significant across the entire wavenumber range. The interference of temperature and bubbles on NIRS is obvious, and the range of influence is concentrated between  $4000\text{--}8000\text{ cm}^{-1}$ . The primary goal of obtaining stable online NIRS data is to reduce the impurity content, maintain a constant sample temperature, and avoid detecting bubbles in the sample.

In summary, this article has designed the online multifunctional NIRS acquisition system with independent intellectual property rights. The overall scheme is shown in Figure 5. The system uses two filters: a stainless-steel filter with a 100-mesh filter inside the extraction tank and a cylindrical filter with a 200-mesh filter. The filters ensure the circulation of the sample while reducing the impurity content. The sample is first cooled by a cooling device in the system and then flows through a water bath temperature control device to ensure that the temperature of the test sample reaches  $35^\circ\text{C} \pm 1^\circ\text{C}$ . The angled design of the water bath temperature control container effectively reduces the number of bubbles during the detection process.

It is worth noting that this article designs a three-channel flow cell that integrates temperature detection, sampling, and NIRS data acquisition, as shown in Figure 5(c). Its characteristics are as follows:

- (1) We used quick-disconnect fittings for easy installation and debugging, and they have strong adaptability in the TCM industry.
- (2) We added a temperature detection window, which detects temperature changes in real-time.
- (3) We added a sampling window, allowing real-time sampling during the detection process, facilitating offline sample collection.
- (4) The three-channel design not only facilitates the circulation of excess liquid but also helps to reduce the flow rate of the detection sample and improve the stability of the detection process.



**Figure 6.** The online NIRS (embedded image) and offline NIRS.

This article applied the developed online multifunctional NIRS acquisition system to the small-scale production line of the XXZOL extraction process and verified its functions in removing impurities, maintaining a constant temperature, and eliminating bubbles. The online NIRS data and collected samples obtained by the system during performance validation are shown in Figure 6.

#### 4.2. Preparation database

This article obtained data for modelling from 16 batches of extraction experiments for XXZOL, as follows:

##### (1) Online NIRS database

We used the system as shown in Section 4.1 to obtain online NIRS data, totalling 2717 spectra. We classified and removed interfering spectra from the online NIRS data and finally obtaining 2412 spectra, forming the online NIRS database.

##### (2) Offline NIRS database

During the manufacturing process, 902 liquid samples were collected using the designed system. Each sample was divided into three parts. The first part was filtered through a filter membrane ( $0.22\text{ }\mu\text{m}$ , PES, China Jinteng) and loaded into a quartz tube with a path length of 2.0 mm. Offline NIRS data were collected using Fourier transform NIRS (MPA II, Bruker, Germany). Each sample is tested three times with a scanning speed of 10 kHz and a resolution of  $8\text{ cm}^{-1}$ . We obtained 902 offline NIRS for training and validation.

##### (3) (Offline) MIRS database

We used a Bruker INVENIOS infrared spectrometer to collect the MIRS from the second part of the sample, taking  $20\text{ }\mu\text{L}$  each time and placing it in the groove of the attenuated

**Table 1.** The CQAs content tested by HPLC (mg/mL<sup>-1</sup>).

| No. | Synephrine | Arecoline | Chlorogenic acid | Forsythoside A | Naringin | Hesperidin | Neohesperidin |
|-----|------------|-----------|------------------|----------------|----------|------------|---------------|
| 1   | 0.0012     | 0.0008    | 0.0003           | 0.0012         | 0.0088   | 0.0023     | 0.0076        |
| 2   | 0.0019     | 0.0013    | 0.0008           | 0.0022         | 0.0126   | 0.0029     | 0.0159        |
| 3   | 0.0024     | 0.0020    | 0.0011           | 0.039          | 0.0238   | 0.0053     | 0.0217        |
| 4   | 0.0033     | 0.0024    | 0.0012           | 0.0051         | 0.0321   | 0.0077     | 0.0293        |
| ... | ...        | ...       | ...              | ...            | ...      | ...        | ...           |
| 219 | 0.0832     | 0.0572    | 0.0152           | 0.5911         | 1.7721   | 1.4180     | 1.7983        |
| 220 | 0.0833     | 0.0570    | 0.0150           | 0.5935         | 1.7687   | 1.4181     | 1.8072        |
| 221 | 0.0833     | 0.0574    | 0.0155           | 0.5970         | 1.7724   | 1.4287     | 1.8088        |
| 222 | 0.0835     | 0.0576    | 0.0157           | 0.5932         | 1.7715   | 1.4235     | 1.8092        |

total reflection attachment for spectral acquisition by transmission. We collect background spectra with air as a reference before measuring each sample. The spectral scanning parameters are 16 scans, a resolution of 4 cm<sup>-1</sup>, a spectral range of 4000–400 cm<sup>-1</sup>. This article obtained 902 (offline) MIRS for training.

#### (4) HPLC database

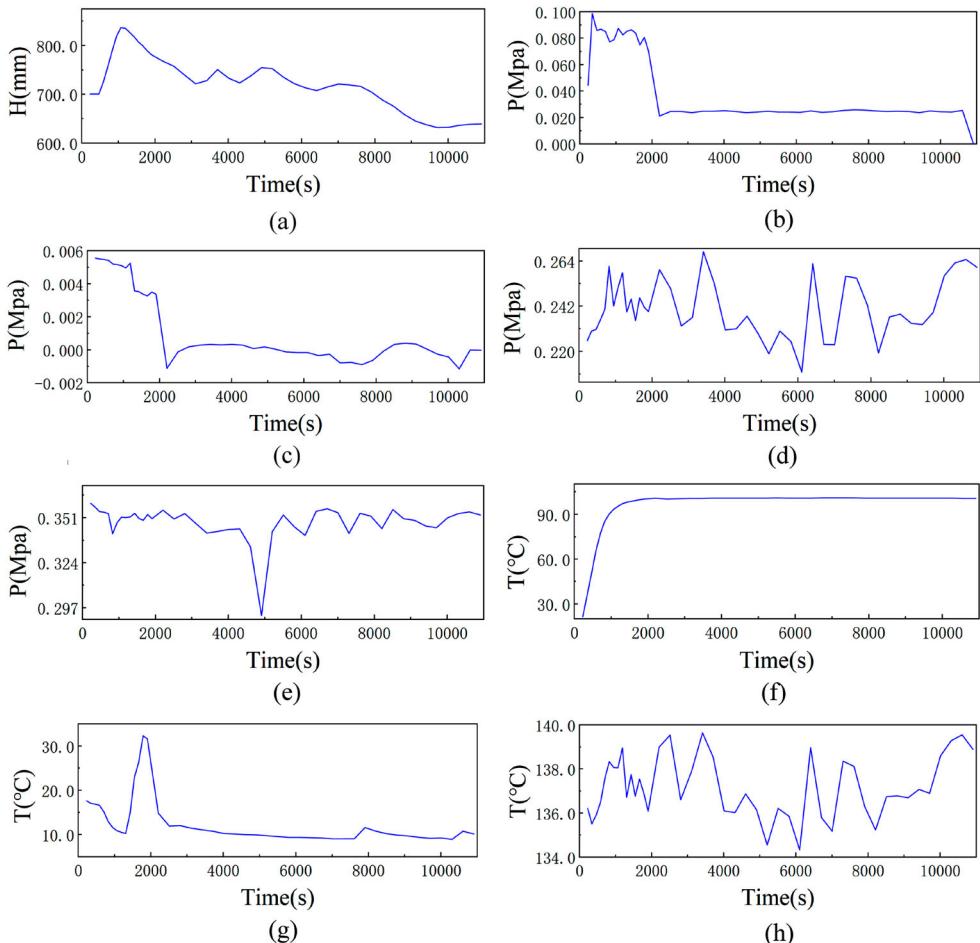
According to the research (J. Zhang et al. 2022), we established the CQAs of XXZOL, which include seven chemical substances: synephrine, arecoline, chlorogenic acid, frysanthoside A, naringin, hesperidin, and neohesperidin. The third part of the sample was used to obtain 222 sets of reference values for CQAs by the HPLC method, as shown in Table 1, which constitutes the HPLC database for model validation.

#### (5) CPPs database

This article used a multifunctional extraction tank and a data acquisition system to record various CPPs data. The data acquisition system was developed by Suzhou Zheyuan Automation Engineering Technology Co., Ltd. The key sensors for process data acquisition included temperature transmitters, pressure transmitters, electromagnetic flow metres, radar level measuring instruments, and others. The experimental data for a certain batch are displayed in the system, as shown in Figure 7. It can be seen that during the extraction process, some parameters varied irregularly within a certain range due to the influence of the factory operating environment, such as the steam header pressure, circulating water pressure, and steam header temperature shown in Figure 7(d,e,h). In addition, sensors may experience sudden data changes due to external interference, as shown in Figure 7(e), where the circulating water pressure suddenly decreases at 5000 s, and in Figure 7(g), where the temperature of the circulating water suddenly increases at the 20,000 s. The entire extraction process can be basically in Figure 7(b): from 0 to 20,000 s, during the heating process, high-temperature steam rapidly heats the medicinal solution to the boiling point; maintain a constant temperature between 2000 and 120,000 s, completing the single extraction process.

### 4.3. Comparative experiments

The proposed framework was developed using Python 3.9 for Windows and aims to assist engineers in evaluating production process quality data. For data processing, especially

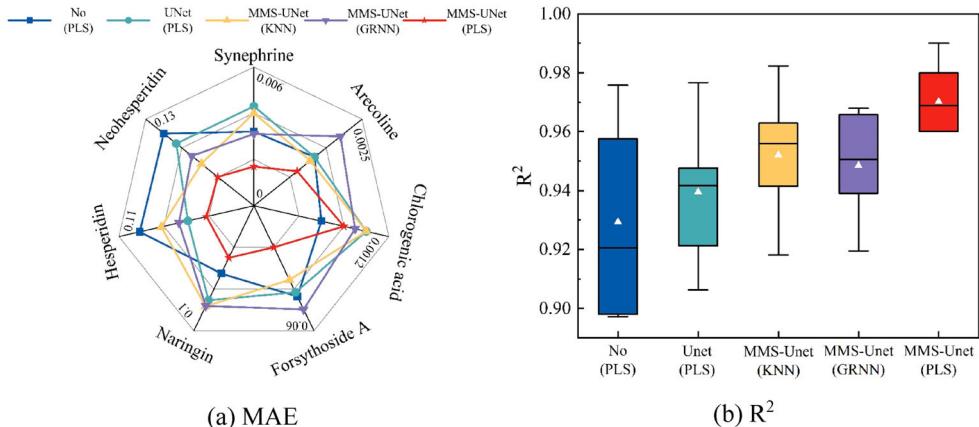


**Figure 7.** The CPPs for the extraction process of a single batch of ZZOXL. (a) The level of extraction liquid, (b) the pressure of vapour, (c) the pressure of extraction tank, (d) the pressure of steam header, (e) the pressure of circulating water, (f) the temperature of extraction liquid, (g) the temperature of circulating water, (h) the temperature of steam header.

for NIRS and MIRS data, we integrated OPUS (version 8.5) and MNE (version 1.3.1). In addition, our CQAs prediction algorithm relies on several other Python libraries: NumPy (version 1.23.5) for numerical calculations, stats models (version 0.13.5) for statistical modelling, and so on.

To verify the performance of MMS-UNet, experiments were conducted to compare it with other algorithms. The comparison algorithms are as follows:

- (1) NO (PLS): The PLS algorithm is directly used for prediction without preprocessing.
- (2) UNet (PLS): The UNet is used for preprocessing, while PLS serves as the final prediction layer. UNet is the original structure, and the loss function based on PDS does not participate in model training.



**Figure 8.** Comparison of different machine learning algorithms for CQAs. (a) MAE (b)  $R^2$ .

- (3) MMS-UNet (KNN): The preprocessing data method is the MMS-UNet framework, and the final prediction layer is K-nearest neighbours (KNN).
- (4) MMS-UNet (GRNN): The final prediction layer of the MMS-UNet framework is replaced by a Generalized Regression Neural Network (GRNN).
- (5) MMS-UNet (PLS): This is the MMS-UNet framework proposed in this chapter, and PLS is used for the final prediction layer.

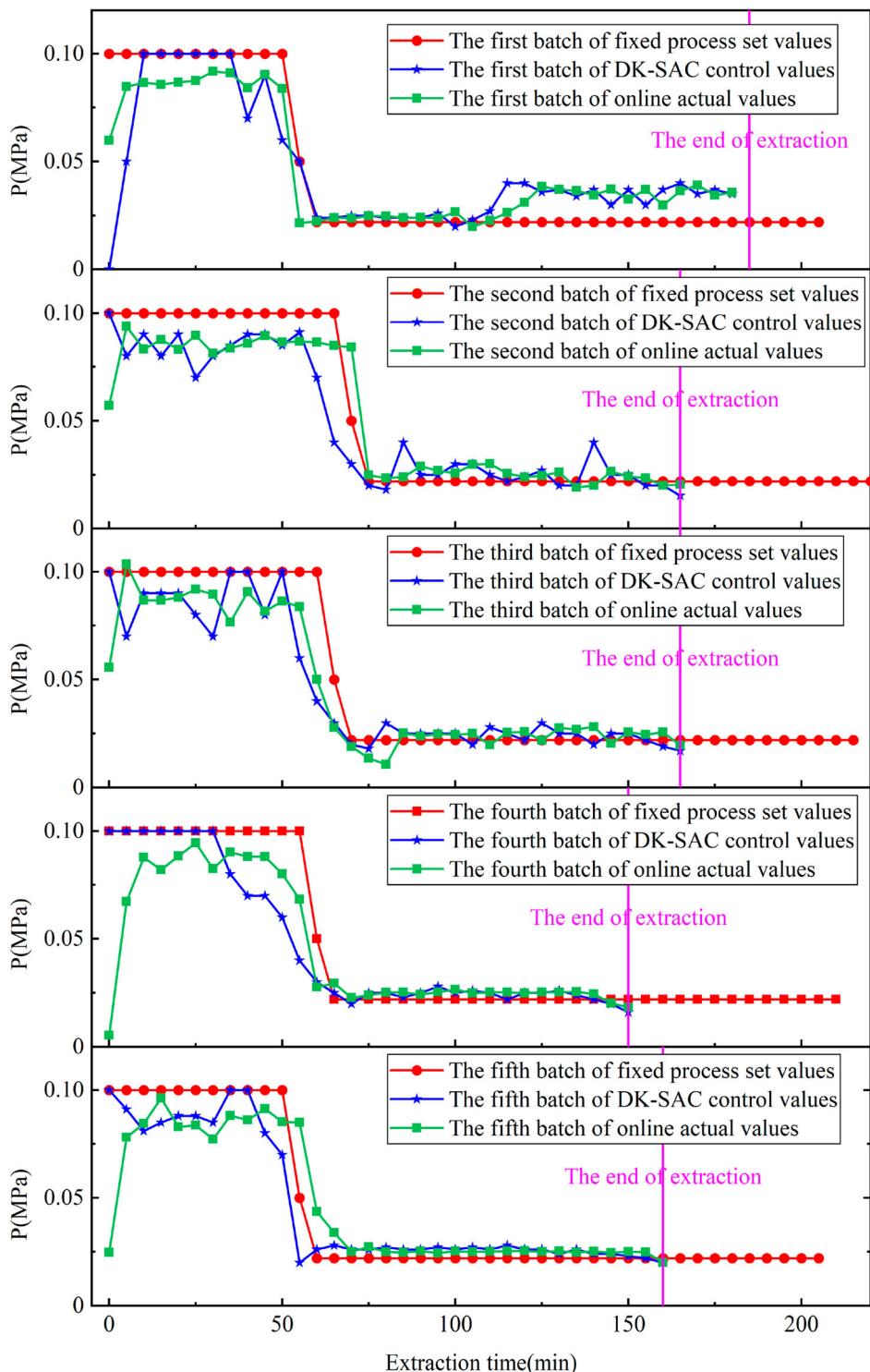
In this section, the training samples for the experiment consist of 902 groups, each group including ZNIRS and LNIRS, with 222 HPLC samples serving as the validation set. The average prediction accuracy of 10 independent experiments is shown in Figure 8.

From Figure 8, it can be seen that compared with other algorithm, the proposed algorithm has improved prediction performance, indicating that ZNIRS contains key information required for modelling. As shown in Figure 8(a), the area enclosed by the red solid line representing MMS-UNet (PLS) is relatively small, indicating that the quantitative model established by MMS-UNet (PLS) for seven CQAs outperforms other machine learning algorithms in terms of MAE.

To verify the effectiveness of DK-SAC, we conducted five batches of XXZOL extraction validation experiments.

The comparison of changes in steam pressure before and after the application of DK-SAC is shown in Figure 9. It can be seen that the boiling time of the extraction process varies for different batches of XXZOL. Among them, the fourth batch takes the longest time, 70 min, while the first and fifth batches take the shortest time, 55 min. The time to reach the extraction endpoint varies among batches, with the fifth batch taking the longest time of 180 min and the second batch taking the shortest time of 150 min.

Table 2 shows the comparison of extraction times before and after the application of DK-SAC. The extraction time before the system application was determined based on the traditional process requirement of maintaining the boiling time for 150 min after the extraction liquid boiled, while the extraction time after the system application was controlled by the DK-SAC as the extraction endpoint time. From the table, it can be seen that the extraction time is greatly reduced after adopting DK-SAC.



**Figure 9.** Comparison of five batches after DK-SAC application.

**Table 2.** Comparison of extraction time after DK-SAC application.

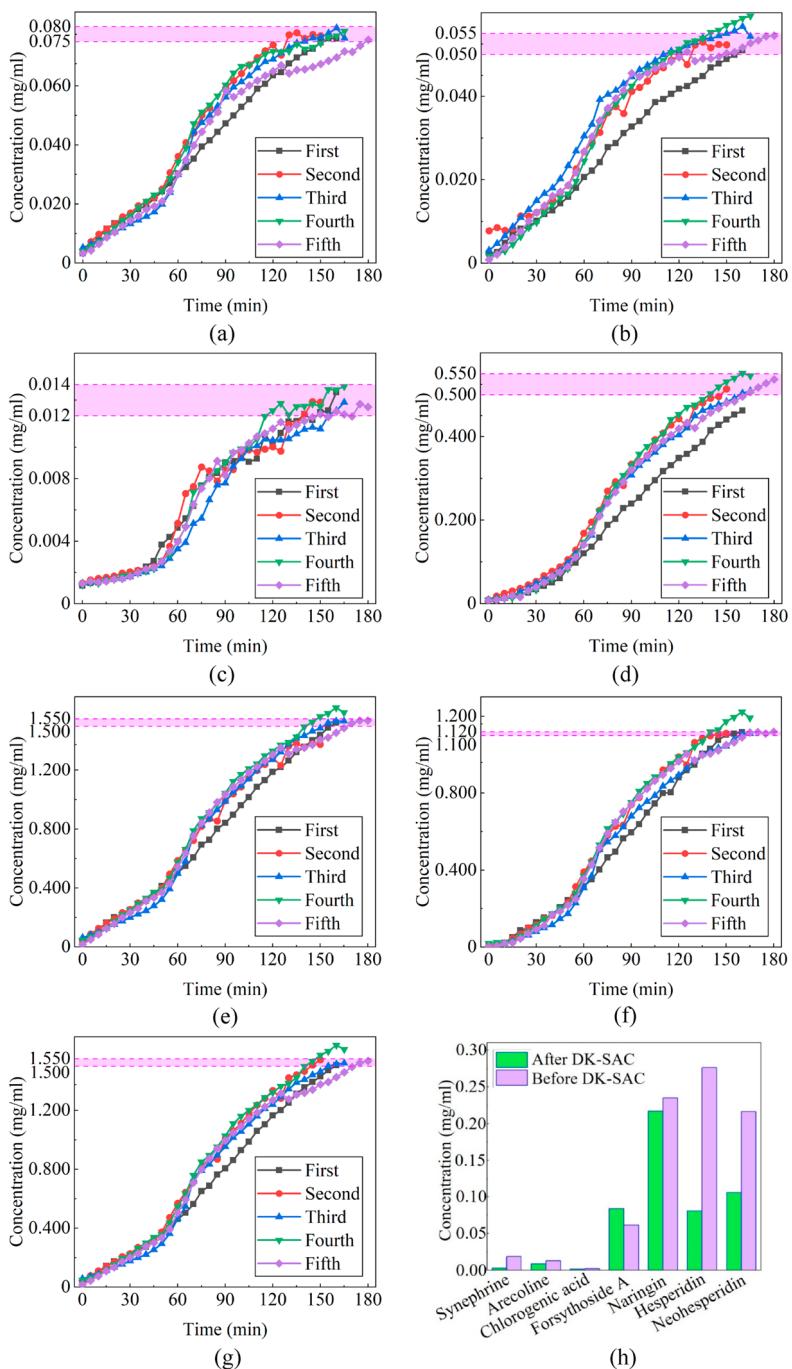
| No. | Extraction time after<br>DK-SAC application (min) | Extraction time before<br>DK-SAC application (min) |
|-----|---|--|
| 1   | 160   | 205  |
| 2   | 150   | 210  |
| 3   | 165   | 215  |
| 4   | 165   | 220  |
| 5   | 180   | 205  |

**Table 3.** Comparison of extraction time after DK-SAC application.

| CQAs             | Batch number | Concentration after<br>DK-SAC application<br>(mg/mL) | concentration before<br>DK-SAC application<br>(mg/mL) |
|------------------|--------------|--|---|
| Synephrine       | 1            | 0.076  | [0.091, 0.072]  |
|                  | 2            | 0.077  |   |
|                  | 3            | 0.076  |   |
|                  | 4            | 0.078  |   |
|                  | 5            | 0.075  |   |
| Arecoline        | 1            | 0.0510   | [0.065, 0.052]  |
|                  | 2            | 0.0523   |   |
|                  | 3            | 0.0543   |   |
|                  | 4            | 0.0593   |   |
|                  | 5            | 0.0545   |   |
| Chlorogenic acid | 1            | 0.0135   | [0.0157, 0.0137]                                      |
|                  | 2            | 0.0129   |   |
|                  | 3            | 0.0129   |   |
|                  | 4            | 0.0139   |   |
|                  | 5            | 0.0126   |   |
| Forsythoside A   | 1            | 0.462  | [0.616, 0.554]  |
|                  | 2            | 0.514  |   |
|                  | 3            | 0.509  |   |
|                  | 4            | 0.546  |   |
|                  | 5            | 0.537  |   |
| Naringin         | 1            | 1.524  | [1.789, 1.555]  |
|                  | 2            | 1.377  |   |
|                  | 3            | 1.537  |   |
|                  | 4            | 1.594  |   |
|                  | 5            | 1.594  |   |
| Hesperidin       | 1            | 1.117  | [1.423, 1.147]  |
|                  | 2            | 1.112  |   |
|                  | 3            | 1.111  |   |
|                  | 4            | 1.192  |   |
|                  | 5            | 1.120  |   |
| Neohesperidin    | 1            | 1.508  | [1.821, 1.605]  |
|                  | 2            | 1.545  |   |
|                  | 3            | 1.521  |   |
|                  | 4            | 1.614  |   |
|                  | 5            | 1.536  |   |

We further analysed the changes of CQAs during the extraction process, and the changes in 7 CQAs from five experimental batches are shown in Figure 10.

Through the analysis of Figure 10(a), it can be seen that at the end of extraction, the synephrine content values of the five batches fall within the pink shadow in the figure, indicating that DK-SAC has achieved precise control over the fluctuation range of synephrine content in different batches. The final synephrine content of the five batches is shown in Table 3.



**Figure 10.** Comparison of 7 CQAs content in five batches. (a) Comparison of synephrine content, (b) comparison of arecoline content, (c) comparison of chlorogenic acid content, (d) comparison of forsythoside A content, (e) comparison of naringin content, (f) comparison of hesperidin content, (g) comparison of neohesperidin content, (h) comparison of CQAs fluctuation intervals.

Other CQAs that have exceeded or not reached the set content range at the extraction endpoint include: in the fourth batch, the content of arecoline, naringin, hesperidin, and neohesperidin exceeded the upper limit of the set content, while in the first batch, the content of folsythoside A did not reach the lower limit of the set content. Due to the different growth trends of each CQA, it is impossible to fully guarantee that all seven CQAs will meet the requirements of the content range. However, through the analysis of XXZOL, it can be concluded that ensuring the CQAs content reaches the set lower limit is more important for the drug's quality. Therefore, the online detection system is effective in controlling the 7 CQAs. Moreover, the contents of CQAs can generally ensure that the set range is reached, indicating that DK-SAC can effectively ensure the consistency of CQAs between batches.

In addition, the changes in 7 CQAs before and after the application of DK-SAC as shown in Table 3, were statistically analysed using data from 16 batch experiments. According to the Pharmacopoeia of the People's Republic of China (2020 edition), the batch with the lowest synephrine content in the experiment still meets the quality requirements. According to Table 3 and Figure 10(h), it can be seen that in the extraction experiments of the five batches, the fluctuation of CQAs content after system application was lower than before, and the fluctuation range was reduced by at least 30%.

## 5. Discussion

In summary, the initially established overall goal (see Section 1) was achieved through the proposed framework combination with its application results from typical case studies. Therefore, by applying implemented methods and systems while utilising expert experience-based knowledge representation, the regulatory and SoSM requirements can be satisfied, and comprehensive improvements in product quality and process design can be achieved. During this process, the following additional issues were resolved. Key data, such as quality, process, and environment, has a decisive impact on design, final quality, efficiency, and is also critical to SoSM. Therefore, it is necessary to obtain real-time information on these data (CQAs) and connect the key link to intelligent control (RQ1). DK-SAC has been proven to have the potential to improve the intelligence of the manufacturing industry, and guidance based on expert knowledge is a necessary condition for quality supervision. Therefore, establishing DK-SAC in TCM manufacturing can effectively upgrade the process design of existing products (RQ2).

As demonstrated by the application case, the proposed system can significantly shorten extraction time and improve production efficiency. By comparing the content of CQAs, it is evident that the quality parameters of all experiments meet the requirements of drug supervision and management, and DK-SAC can stabilise the CQAs content range, improving consistency between batches in the production process. The proposed framework can consider specific quality information in advance and incorporate it into the process design environment, thus achieving production designs with quality consistency. For specific and strict quality control requirements in production, a certain degree of self-adaptive control can be achieved. Therefore, the proposed framework has great potential to assist production industries with limited regulations or industry rules, enabling them to rethink existing designs and ultimately consider data-knowledge hybrid driven control strategies.

## 6. Conclusion and outlook

The flexibility of information technology and the need for adaptive control provide the opportunity to restructure traditional process control designs through data – knowledge hybrid driven approaches. However, this requires incorporating expert knowledge and the most important quality data into the newly conceived control framework. Therefore, this article focuses on obtaining key data through a comprehensive approach using hardware and algorithms to ensure the quality of TCM manufacturing. This study utilises the fusion of multi-source data (NIRS, MIRS, and HPLC) to facilitate more accurate acquisition of CQAs, achieving efficient and accurate information representation and transmission. By analysing data and considering quality information as a prerequisite, the adaptability of the overall control system will be ensured. The case study of XXZOL demonstrates the workflow of the system, thereby demonstrating the effectiveness of the proposed DK-SAC.

Although the introduced framework is developed for the extraction process of TCM, it can also be applied to other TCM processes and even the entire pharmaceutical industry. For this purpose, the general method can be retained by determining the design criteria and multi-source data fusion methods for the corresponding online CQAs data acquisition system and integrating them into the ontology of the data-knowledge hybrid driven adaptive control system after normalisation. In addition, we plan to explore online detection paradigms for different forms of raw materials, such as solids and powders, to provide more universal detection tools for pharmaceutical engineering.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This work was supported by the National Natural Science Foundation of China under Grant 52405566, and the Fundamental Research Program of Shanxi Province under Grant 202403021222177, 202403021222145, and the Key Laboratory of Computing Power Network and Information Security open project under Grant 2023PY009, and the Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi under Grant 2024L182.

## ORCID

Zhongxin Chen  <http://orcid.org/0000-0001-5757-2972>

## References

- An, H., C. Zhai, F. Zhang, Q. Ma, J. Sun, Y. Tang, and W. Wang. 2023. "Quantitative Analysis of Chinese Steamed Bread Staling Using NIR, MIR, and Raman Spectral Data Fusion." *Food Chemistry* 405:134821. <https://doi.org/10.1016/j.foodchem.2022.134821>
- Brooks, S., and R. Roy. 2021. "An Overview of Self-Engineering Systems." *Journal of Engineering Design* 32 (8): 397–447. <https://doi.org/10.1080/09544828.2021.1914323>
- Chen, Z., Y. Shen, B. Chen, J. Zhou, P. Huang, H. Zang, and Y. Guan. 2023. "Vaseline: Realize Online Detection and Augmented NIR Using Deep Learning." *Engineering Applications of Artificial Intelligence* 126:106684. <https://doi.org/10.1016/j.engappai.2023.106684>
- Chen, Z., Y. Tang, Z. Gao, J. Zhou, and P. Huang. 2022. "An Ensemble Active Learning for a Fluidized Bed Granulation in the Pharmaceutical Industry." *Journal of Process Control* 118:16–25. <https://doi.org/10.1016/j.jprocont.2022.08.007>

- Deidda, R., P.-Y. Sacre, M. Clavaud, L. Coic, H. Avohou, P. Hubert, and E. Ziemons. **2019**. "Vibrational Spectroscopy in Analysis of Pharmaceuticals: Critical Review of Innovative Portable and Handheld NIR and Raman Spectrophotometers." *TrAC Trends in Analytical Chemistry* 114:251–259. <https://doi.org/10.1016/j.trac.2019.02.035>
- Dong, X., H. Ding, D. Gao, G. Zheng, J. Wang, and Q. Lang. **2025**. "Multi-source Fault Data Fusion Diagnosis Method Based on Hyper-Feature Space Graph Collaborative Embedding." *Advanced Engineering Informatics* 64:103092. <https://doi.org/10.1016/j.aei.2024.103092>
- Du, J., Z. Huang, C. Li, and L. Jiang. **2024**. "Quantitative Analysis of the Illegal Addition of Atenolol in Panax Notoginseng Based on NIR-MIR Spectral Data Fusion and Calibration Transfer." *Rsc Advances* 14 (18): 12428–12437. <https://doi.org/10.1039/D3RA08183D>
- Gan, F., and J. Luo. **2023**. "Simple Dilated Convolutional Neural Network for Quantitative Modeling Based on Near Infrared Spectroscopy Techniques." *Chemometrics and Intelligent Laboratory Systems* 232:104710. <https://doi.org/10.1016/j.chemolab.2022.104710>
- Gao, L., L. Zhong, Y. Wei, L. Li, A. Wu, L. Nie, J. Yue, et al. **2023**. "A new Perspective in Understanding the Processing Mechanisms of Traditional Chinese Medicine by Near-Infrared Spectroscopy with Aquaphotomics." *Journal of Molecular Structure* 1284:135401. <https://doi.org/10.1016/j.molstruc.2023.135401>
- Garcia-Munoz, S., and J. Mercado. **2013**. "Optimal Selection of raw Materials for Pharmaceutical Drug Product Design and Manufacture Using Mixed Integer Nonlinear Programming and Multivariate Latent Variable Regression Models." *Industrial & Engineering Chemistry Research* 52 (17): 5934–5942. <https://doi.org/10.1021/ie3031828>
- Guo, T., W.-H. Feng, X.-Q. Liu, H.-M. Gao, Z.-M. Wang, and L.-L. Gao. **2016**. "Fourier Transform mid-Infrared Spectroscopy (FT-MIR) Combined with Chemometrics for Quantitative Analysis of Dex-trin in Danshen (*Salvia miltiorrhiza*) Granule." *Journal of Pharmaceutical and Biomedical Analysis* 123:16–23. <https://doi.org/10.1016/j.jpba.2015.11.021>
- Hao, N., J. Ping, X. Wang, X. Sha, Y. Wang, P. Miao, C. Liu, and W. Li. **2024**. "Data Fusion of Near-Infrared and mid-Infrared Spectroscopy for Rapid Origin Identification and Quality Evaluation of *Lonicerae japonicae Flos*." *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 320:124590. <https://doi.org/10.1016/j.saa.2024.124590>
- Hu, X., M. A. Naiel, A. Wong, M. Lamm, and P. Fieguth. **2019**. "RUNet: A Robust UNet Architecture for Image Super-Resolution." *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Long Beach, CA, USA, 505–507. <https://doi.org/10.1109/CVPRW.2019.00073>
- Jiao, R., J. Luo, J. Malmqvist, and J. Summers. **2022**. "New Design: Opportunities for Engineering Design in an era of Digital Transformation." *Journal of Engineering Design* 33 (10): 685–690. <https://doi.org/10.1080/09544828.2022.2147270>
- Kang, C., G. Zhang, G. Mu, H. Guo, T. Yuan, C. Zhao, X. Li, and Q. Zhang. **2024**. "Solar-heat Pump Com-bined Drying with Phase Change Heat Storage: Multi-Energy Self-Adaptive Control." *Renewable Energy* 230:120867. <https://doi.org/10.1016/j.renene.2024.120867>
- Keung, K. L., C. K. M. Lee, T.-T. Tsang, C. Liu, and I. Misbah. **2024**. "The Impact of Customisation Expe-rience and co-Design Value on the 3D Printed Specimen." *Journal of Engineering Design* 35 (9): 1153–1184. <https://doi.org/10.1080/09544828.2024.2373037>
- Lan, Z., Y. Zhang, Y. Sun, D. Ji, S. Wang, T. Lu, H. Cao, and J. Meng. **2020**. "A mid-Level Data Fusion Approach for Evaluating the Internal and External Changes Determined by FT-NIR, Electronic Nose and Colorimeter in *Curcumae rhizoma* Processing." *Journal of Pharmaceutical and Biomedical Analysis* 188:113387. <https://doi.org/10.1016/j.jpba.2020.113387>
- Lei, W., X. Li, S. Li, F. Zhou, Y. Guo, M. Zhang, X. Jin, and H. Zhang. **2024**. "Targeting Neutrophils Extra-cellular Traps, a Promising Anti-Thrombotic Therapy for Natural Products from Traditional Chinese Herbal Medicine." *Biomedicine & Pharmacotherapy* 179:117310. <https://doi.org/10.1016/j.biopharm.2024.117310>
- Lei, T., and D.-W. Sun. **2020**. "A Novel NIR Spectral Calibration Method: Sparse Coefficients Wavelength Selection and Regression (SCWR)." *Analytica Chimica Acta* 1110:169–180. <https://doi.org/10.1016/j.aca.2020.03.007>

- Liu, N., C. Liu, L. Chen, J. Yu, X. Sun, S. Zhang, and J. Wu. 2024. "Research on Model Transfer Strategies Based on the Fusion of NIR-MIR Spectral Data." *Infrared Physics & Technology* 142:105545. <https://doi.org/10.1016/j.infrared.2024.105545>
- Liu, L., Z.-T. Zuo, Y.-Z. Wang, and F.-R. Xu. 2020. "A Fast Multi-Source Information Fusion Strategy Based on FTIR Spectroscopy for Geographical Authentication of Wild *Gentiana rigescens*." *Microchemical Journal* 159:105360. <https://doi.org/10.1016/j.microc.2020.105360>
- Lukacs, M., G. Bazar, B. Pollner, R. Henn, C. G. Kirchler, C. W. Huck, and Z. Kovacs. 2018. "Near Infrared Spectroscopy as an Alternative Quick Method for Simultaneous Detection of Multiple Adulterants in Whey Protein-Based Sports Supplement." *Food Control* 94:331–340. <https://doi.org/10.1016/j.foodcont.2018.07.004>
- Ma, C., L. Ma, Z. Wang, N. Li, M. Li, J. Wang, X. Wang, et al. 2023. "Original End-to-End Smart Diagnosis Framework of Systematic Critical Quality Attributes Meets FDA Standards of Phytotherapy by Biosensor and Multi-Information Fusion Coupled with AI Algorithm." *Green Chemistry* 25 (1): 384–398. <https://doi.org/10.1039/D2GC03835H>
- Mao, S., Q.-Y. Du, M. He, L. Sun, J. Shi, X. Zhou, X.-Z. Zhu, Y.-J. Yu, and X. Zhang. 2024. "A Strategy of Q-Markers Identification Based on Effect, Property Flavour Material Basis and Rapid Quantitative Evaluation via Near-Infrared Spectroscopy and Chemometric Methods for the Quality Control of *Flos Trollii* (Ft)." *Journal of Ethnopharmacology* 337:118883. <https://doi.org/10.1016/j.jep.2024.118883>
- Nakata, K., R. Orihara, Y. Mizuoka, and K. Takagi. 2017. "A Comprehensive Big-Data-Based Monitoring System for Yield Enhancement in Semiconductor Manufacturing." *Ieee Transactions on Semiconductor Manufacturing* 30 (4): 339–344. <https://doi.org/10.1109/TSM.2017.2753251>
- Prikerznik, M., and S. Srcic. 2021. "Multivariate Analysis for Optimization and Validation of the Industrial Tablet-Manufacturing Process." *Drug Development and Industrial Pharmacy* 47 (1): 61–71. <https://doi.org/10.1080/03639045.2020.1851244>
- Radcliffe, A. J., and G. V. Reklaitis. 2021. "Process Monitoring under Uncertainty: An Opportunity for Bayesian Multilevel Modelling." In *Computer Aided Chemical Engineering*, edited by M. Türkay, and R. Gani, 1377–1382. Amsterdam, Netherlands: Elsevier.
- Robert, G., and R. Gosselin. 2022. "Evaluating the Impact of NIR pre-Processing Methods via Multiblock Partial Least-Squares." *Analytica Chimica Acta* 1189:339255. <https://doi.org/10.1016/j.aca.2021.339255>
- Siiskonen, M., R. Govender, J. Malmqvist, and S. Folestad. 2023. "Modelling the Cost-Benefit Impact of Integrated Product Modularisation and Postponement in the Supply Chain for Pharmaceutical Mass Customisation." *Journal of Engineering Design* 34 (10): 865–896. <https://doi.org/10.1080/09544828.2023.2266330>
- Singh, R., A. Sahay, K. M. Karry, F. Muzzio, M. Ierapetritou, and R. Ramachandran. 2014. "Implementation of an Advanced Hybrid MPC-PID Control System Using pat Tools Into a Direct Compaction Continuous Pharmaceutical Tablet Manufacturing Pilot Plant." *International Journal of Pharmaceutics* 473 (1-2): 38–54. <https://doi.org/10.1016/j.ijpharm.2014.06.045>
- Sun, L., J. Xi, Q. Xia, Z. Li, and A. Kumail. 2019. "Optimal Control Strategy of Multi-Stage Pharmaceutical." *IFAC-PapersOnLine* 52 (10): 370–375. <https://doi.org/10.1016/j.ifacol.2019.10.059>
- Tao, L., B. Via, Y. Wu, W. Xiao, and X. Liu. 2019. "NIR and MIR Spectral Data Fusion for Rapid Detection of *Lonicera japonica* and *Artemisia annua* by Liquid Extraction Process." *Vibrational Spectroscopy* 102:31–38. <https://doi.org/10.1016/j.vibspec.2019.03.005>
- Wang, F., Z. Jiang, K. Cheng, G. Fan, and H. Zhou. 2024. "A Multimodal Approach for Fostering Knowledge Transfer in Engineering Design Activity: An eye-Tracking and fNIRS Study." *Journal of Engineering Design* 35 (12): 1518–1549. <https://doi.org/10.1080/09544828.2024.2380621>
- Wang, K., Z. Wang, H. Xu, Z. Lan, X. Lin, J. Ren, and S. Kong. 2025. "Experimental Study on Online Detection of Near-Infrared Spectroscopy Suitable for Continuous Drug Production." *Journal of Drug Delivery Science and Technology* 104:106528. <https://doi.org/10.1016/j.jddst.2024.106528>
- Wang, Q., N. Wei, L. Xu, H. Hua, J. Li, Y. Jiang, and L. Chen. 2023. "TCM Fingerprint Database: A Digital Approach to Scientifically Reflect the Internal Quality of Traditional Chinese Medicine." *Pharmacological Research - Modern Chinese Medicine* 7:100261. <https://doi.org/10.1016/j.prmcm.2023.100261>
- Wang, A., P. Yang, J. Chen, Z. Wu, Y. Jia, C. Ma, and X. Zhan. 2019. "A new Calibration Model Transferring Strategy Maintaining the Predictive Abilities of NIR Multivariate Calibration Model

- Applied in Different Batches Process of Extraction." *Infrared Physics & Technology* 103:103046. <https://doi.org/10.1016/j.infrared.2019.103046>
- Yenduri, G., A. P. Costa, X. Xu, and D. J. Burgess. 2022. "Impact of Critical Process Parameters and Critical Material Attributes on the Critical Quality Attributes of Liposomal Formulations Prepared Using Continuous Processing." *International Journal of Pharmaceutics* 619:121700. <https://doi.org/10.1016/j.ijpharm.2022.121700>
- Zhang, Q., J. Liu, L. Li, X. Chen, and R. Wang. 2024. "Automatic Generation of System Model Diagrams Driven by Multi-Source Heterogeneous Data." *Journal of Engineering Design* 35 (11): 1442–1486. <https://doi.org/10.1080/09544828.2024.2360853>
- Zhang, J.-B., B. Wang, Y.-F. Zhang, Y. Wu, M.-X. Li, T. Gao, T.-L. Lu, Z.-H. Bian, and L.-L. Su. 2024. "E-eye and FT-NIR Combined with Multivariate Algorithms to Rapidly Evaluate the Dynamic Changes in the Quality of *Gastrodia elata* During Steaming Process." *Food Chemistry* 439:138148. <https://doi.org/10.1016/j.foodchem.2023.138148>
- Zhang, J., X. Xu, L. Li, H. Li, L. Gao, X. Yuan, H. Du, Y. Guan, and H. Zang. 2022. "Multi Critical Quality Attributes Monitoring of Chinese Oral Liquid Extraction Process with a Spectral Sensor Fusion Strategy." *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 278:121317. <https://doi.org/10.1016/j.saa.2022.121317>
- Zhang, T., and X. Zhang. 2022. "Squeeze-and-excitation Laplacian Pyramid Network with Dual-Polarization Feature Fusion for Ship Classification in SAR Images." *IEEE Geoscience and Remote Sensing Letters* 19:1–5.
- Zhou, T., X. Ming, T. Han, Y. Bao, X. Liao, Q. Tong, S. Liu, H. Guan, and Z. Chen. 2023. "Smart Experience-Oriented Customer Requirement Analysis for Smart Product Service System: A Novel Hesitant Fuzzy Linguistic Cloud DEMATEL Method." *Advanced Engineering Informatics* 56:101917. <https://doi.org/10.1016/j.aei.2023.101917>
- Zhu, W., P. Ouyang, and M. Kong. 2024. "Research on the Evolution Mechanism of Intelligent Manufacturing Transformation of Chinese Pharmaceutical Manufacturing Enterprises Based on System Dynamics." *Heliyon* 10 (13): e33959. <https://doi.org/10.1016/j.heliyon.2024.e33959>