Analytical-Chemistry-Informed Transformer for Infrared Spectra Modeling

Shiluo Huang[1], Yining Jin[2], Wei Jin[3*], Ying Mu[3]

[1] School of Computing and Artificial Intelligence, Southwestern University of Finance and Economics, Chengdu, China
[2] Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Alberta, Canada
[3] Research Centre for Analytical Instrumentation, State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou, China
huangsl@swufe.edu.cn, yining.jin.academic@gmail.com, {jinweimy, muying}@zju.edu.cn

Abstract

Infrared (IR) spectroscopy is a fundamental technique in analytical chemistry. Recently, deep learning (DL) has drawn great interest as the modeling method of infrared spectral data. However, unlike vision or language tasks, IR spectral data modeling is faced with the problem of calibration transfer and has distinctive characteristics. Introducing the prior knowledge of IR spectroscopy could guide the DL methods to learn representations aligned with the domain-invariant characteristics of spectra, and thus improve the performance. Despite such potential, there is a notable absence of DL methods that incorporate such inductive bias. To this end, we propose Analytical-Chemistry-Informed Transformer (ACT) with two modules informed by the field knowledge in analytical chemistry. First, ACT includes learnable spectral processing inspired by chemometrics, which comprises spectral pre-processing, tokenization, and post-processing. Second, a straightforward yet effective representation learning mechanism, namely spectral-attention, is incorporated into ACT. Spectral-attention utilizes the intra-spectral and inter-spectral correlations to extract spectral representations. Empirical results show that ACT has achieved competitive results in 9 analytical tasks covering applications across pharmacy, chemistry, and agriculture. Compared with existing networks, ACT reduces the root mean square error of prediction (RMSEP) by more than 20% in calibration transfer tasks. These results indicate that DL methods in IR spectroscopy could benefit from the integration of prior knowledge in analytical chemistry.

Code — https://github.com/Siryokait/ACT4IRSpectra

Introduction

Infrared (IR) spectroscopy plays a crucial role in both scientific research (Kalashnikov et al. 2016; Ramzan et al. 2024; Bredenbeck et al. 2008) and industrial application (Porep, Kammerer, and Carle 2015; Yu et al. 2021a; Miao et al. 2023), providing a rapid, non-destructive and economical solution to chemical analysis (Weber et al. 2023; Martens et al. 2017). As vibrational spectroscopy, IR spectroscopy studies the absorbance of light resulting from molecular vibrations, which manifests itself as spectra containing peaks or overtones at different wavelengths (Mishra et al. 2022a). IR spectroscopy has been introduced to the analysis of complex samples in biological, medical, and chemical applications with the assistance of chemometrics (Åsmund Rinnan, van den Berg, and Engelsen 2009). IR spectra could reveal the categories and specific contents of samples, which is usually realized by a calibration model (Manley 2014).

Despite the success of chemometric methods, it is still challenging to establish a robust calibration model across capricious spectra-collecting environments (Yang et al. 2024). It is recognized that the performance of calibration models will degrade when handling the data collected with different spectrometers, varied sampling protocols, or different environmental conditions (Nikzad-Langerodi et al. 2018). Illustrated in Fig. 1, calibration transfer aims to improve the prediction ability under such a condition, where training (source) data and testing (target) data might not follow the i.i.d. assumption due to different spectra-collecting processes. Conventional calibration transfer studies can be categorized as domain adaptation methods (Vettoruzzo et al. 2024), which either retrain (Yu et al. 2021b; Mishra et al. 2021) or regularize (Nikzad-Langerodi et al. 2018; Zhang, Zhou, and Li 2023) the calibration model with the data from the target domain. However, collecting extra data from the target domain could be costly and such data collection is not always practical.

In the last decade, deep learning (DL) has dramatically improved the state-of-the-art various fields (LeCun, Bengio, and Hinton 2015; Vaswani et al. 2017; Wu et al. 2021; Liu et al. 2024; Nie et al. 2022). On the contrary, DL for IR spectroscopy is still in its infancy (Mishra et al. 2022b). Besides the lack of large open datasets, the absence of specialized DL architectures may be responsible for this delay. Convolutional networks (CNNs), Transformers, and other neural networks have been introduced to IR spectra modeling (Zhang et al. 2019; Jung, Jung, and Cole 2023; Enders et al. 2021). Recently, researchers have tried to boost the performance via data augmentation and learn the augmented data with CNNs (Wang et al. 2023; Wu et al. 2024; Chen, Zhou, and Ren 2024), which achieves competitive performance. Although these networks alleviate the reliance on spectral pre-processing (Mishra et al. 2022b), the learned models still suffer performance degradation when training data and testing data are generated from different spectral-collecting processes. Meanwhile, existing networks seldom consider the

*Corresponding author

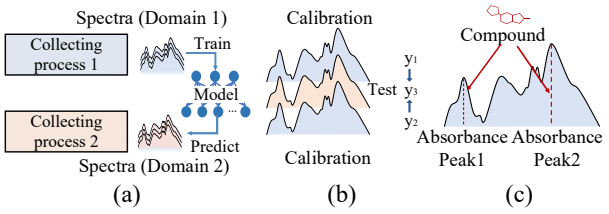

Figure 1: Characteristics for IR spectroscopy: (a) calibration transfer: the performance of models may drastically decline when the testing samples are collected using a different spectrometer, even if it is from the same manufacturer; (b) quantitative meaning of IR spectra relies on labeled calibration spectra; (c) non-adjacent absorbance peaks resulting from the same chemical compound exist.

characteristics of IR spectra and the insights provided by chemometrics.

Incorporating prior knowledge of chemometrics and spectral characteristics into the deep networks is a potential solution. The prior knowledge could reveal the intrinsic meaning of IR spectra. The characteristics of IR spectral data differ from the characteristics of vision and language data. As shown in Fig. 1, both inter- and intra-spectral correlations are vital for IR spectra analysis. IR spectroscopy is an indirect method and thus inter-spectral correlations determine the quantitative meaning of a single spectrum (Mishra et al. 2022b). A compound with multiple chemical groups could absorb infrared radiation at multiple wavelengths (Mishra et al. 2022a), and the intra-spectral correlation therefore contains the information of possible chemical compounds. The absorbance peaks caused by a single compound are usually non-adjacent, also resulting in long-range dependencies along the spectral axis. Incorporating such field knowledge would help the network to learn domain-invariant representations without the need for data in the target domain, realizing domain generalization (Vettoruzzo et al. 2024) in IR spectroscopy.

However, there is a lack of deep networks incorporating the prior knowledge of chemometrics and spectral characteristics. To address this issue, a Transformer specialized in IR spectroscopy is proposed in this paper, namely Analytical-Chemistry-Informed Transformer (ACT). Based on Vanilla Transformer (Vaswani et al. 2017), ACT implements several modifications in terms of spectral processing and representation learning. Firstly, a learnable spectral processing module is proposed and incorporated into ACT, which is an integration of pre-processing, tokenization, and post-processing. Following the insights provided by chemometric baseline correction, we propose reversible spectral pre-processing. Secondly, ACT introduces a spectral-attention mechanism in place of self-attention. Aligned with the prior knowledge in IR spectroscopy, spectral-attention is designed to exploit the similarities among spectra and the correlations among spectral bands. Guided by field knowledge, ACT could focus on proper spectral peaks like human experts and learn domain-invariant representations. The contributions are summarized as follows:

1. We propose ACT, a Transformer specialized in IR spectral data modeling, to cope with the calibration transfer tasks by learning domain-invariant representations.
2. Inspired by conventional chemometrics, we propose learnable spectral processing integrated with ACT, which incorporates reversible baseline correction.
3. Following the prior knowledge of IR spectroscopy, we propose the spectral-attention mechanism that considers both the inter- and intra-spectral correlations.
4. We conduct comprehensive experiments with 9 IR spectra modeling tasks in pharmacy, chemistry, and agriculture, where ACT outperforms the comparison methods in terms of both calibration transfer and regular analysis.

Preliminary and Related Work

Infrared Spectroscopy and Chemometrics IR spectroscopy can be generally divided into three categories: the near-infrared (NIR), the mid-infrared (MIR), and the far-infrared (Debus et al. 2021). Among these categories, NIR and MIR spectroscopy have widespread applications in pharmacy (Li et al. 2014), chemistry (Miao et al. 2023), agriculture (Porep, Kammerer, and Carle 2015), biology (Ramzan et al. 2024) and material science (John-Herpin et al. 2023). As an indirect method, IR spectroscopy requires a calibration model for qualitative and quantitative analysis. Chemometrics studies the modeling methods for chemical data including IR spectra, where machine learning techniques have a crucial role (Marini 2009). In the conventional chemometric modeling pipeline, knowledge-driven spectral pre-processing (feature engineering) is usually essential (Mishra et al. 2022c; Mokari, Guo, and Bocklitz 2023). However, considerable error could exist in the estimated baselines of these pre-processing methods leading to 'over-processing', which results in the loss of useful information. Meanwhile, the analysis targets may also correlate with physical characteristics that contribute to baselines. In this paper, the proposed ACT incorporates a unique learnable pre-processing module with baseline reconstruction to reclaim valuable information from the estimated baselines.

Deep Learning for Infrared Spectra Modeling As DL bloomed in the last decade, deep networks have been utilized for IR spectral data modeling (Wang et al. 2022a; Cui and Fearn 2018). CNN along with its variants is one of the most popular deep networks in the field of infrared spectroscopy. DeepSpectra (Zhang et al. 2019), one of the earliest deep spectral modeling methods, utilizes the structure of Inception (Szegedy et al. 2016). Transformer also draws considerable attention (Chen, Zhou, and Ren 2024; Wang et al. 2022b). Spectraformer (Chen, Zhou, and Ren 2024) combines the encoder of Vanilla Transformer with multi-layer CNN. The above studies adopt existing network structures without modifications specific to IR spectra (Chen and Wang 2019). Recent studies like AggMapNet (Wang et al. 2023) and TeaNet (Wu et al. 2024) have introduced specialized data augmentation methods to CNNs. However, there is still a lack of deep networks that consider the field knowledge of IR spectroscopy and the insights provided by chemometric methods.

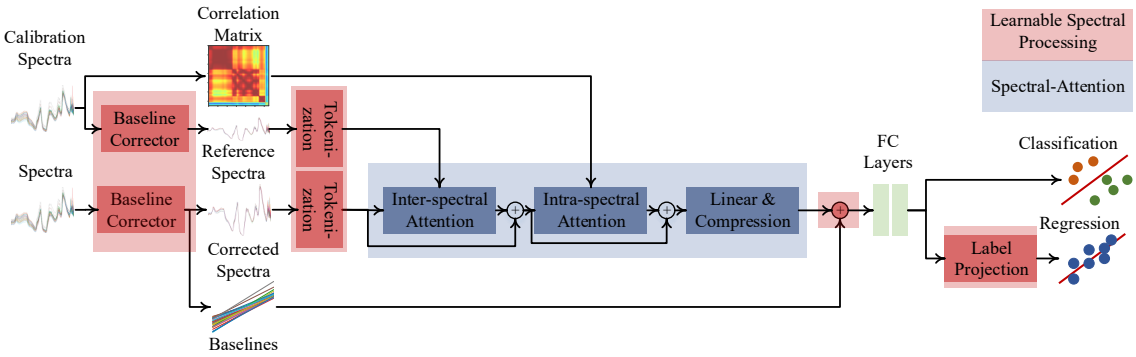

Figure 2: ACT architecture. The learnable spectral processing module (red blocks) first splits raw spectra into baselines and corrected spectra, then recovers the baseline information, and finally adjusts predictions for regression tasks. Spectral-attention (blue blocks) utilizes reference spectra and correlation matrix to refine the attention map. The challenges/characteristics along with corresponding modules are also presented at the bottom left of the figure.

Calibration Transfer A well-known problem with IR spectroscopy is that the predictions of a calibration model are reliable only if the calibrating (training) and testing spectra are collected with the same spectral-collecting process (Nikzad-Langerodi et al. 2018). Differences in measurement environments, instruments, and sample-handling protocols could disturb the calibration model (Varmuza and Filzmoser 2016). For instance, the model's performance may drastically decline when the testing samples are collected using a different spectrometer, even if it is from the same manufacturer. Both the conventional methods and the DL methods suffer a performance degradation during tackling data from a different spectra-collecting process (Mishra and Passos 2021). Researchers try to alleviate this problem by retraining or regularizing the model with the data from target domain, namely data within the same domain as the testing data (Yu et al. 2021b; Mishra et al. 2021; Nikzad-Langerodi et al. 2018; Zhang, Zhou, and Li 2023). However, collecting extra data could be costly and the data from target domain is not always available. A generic model that could achieve satisfactory performance across different spectra-collecting processes is therefore attractive. To this end, ACT tries to incorporate domain-invariant knowledge to improve performance without relying on data from the target domain.

Proposed Method

ACT incorporates two specialized designs: learnable spectral processing and spectral-attention, which is illustrated in Fig. 2. In this paper, we denote a set of IR spectra as $\boldsymbol{X} = \{\boldsymbol{x}_i\}_{i=1}^{n}$, while the corresponding labels (analysis targets) are denoted as $\boldsymbol{Y}$. Given a set of calibration spectra and corresponding labels $\{\boldsymbol{X}_{cal}, \boldsymbol{Y}_{cal}\}$, ACT aims to analyze the testing spectra $\boldsymbol{X}_{test}$ and predict the labels $\boldsymbol{Y}_{test}$, where $\boldsymbol{X}_{test}$ might be collected with different instruments, sample preparing protocols, etc. We use n and b to represent the number of samples and spectral bands respectively.

Learnable Spectral Processing

Learnable spectral processing incorporates a reversible baseline corrector, which utilizes the widely used corrector (denoted as BaseLineCorr) based on iterative polynomial fitting for baseline estimation (Gan, Ruan, and Mo 2006). The corrector assumes that baselines within chemical signals can be estimated by a polynomial with lower power. The spectra are corrected in two steps: first estimating and then removing baselines.

Distinguishing from conventional spectral pre-processing (Peng et al. 2011), the reversible baseline correction keeps baselines and reverses the correction in the subsequent modules. For spectral data $\boldsymbol{X} \in \mathbb{R}^{n\times b}$, the baseline correction process is denoted as:

$$\begin{aligned} \boldsymbol{\mathcal{B}} &= \mathrm{BaseLineCorr}(\boldsymbol{X}) \\ \boldsymbol{\mathcal{X}} &= \boldsymbol{X} - \boldsymbol{\mathcal{B}}, \end{aligned} \tag{1}$$

where $\boldsymbol{\mathcal{B}}, \boldsymbol{\mathcal{X}} \in \mathbb{R}^{n\times b}$ denote the baselines and corrected spectra respectively. $\boldsymbol{\mathcal{X}}$ is subsequently normalized. For each spectrum $\S_i \in \boldsymbol{\mathcal{X}}$, the varied peak shape leads to different analytical meanings of intensities at a single wavelength. As a result, the same absorbance intensities could represent different levels of chemical groups/compounds. It is hard to represent such neighboring information by utilizing the intensity values at each position exclusively.

To preserve the local spectral shape, the corrected spectral data $\boldsymbol{\mathcal{X}}$ are divided into overlapped patches. The patch at a single wavelength consists of the spectral intensities within a neighboring window. For a single spectral point at wavelength λ_j, l neighboring spectral points at higher wavelengths and l neighboring spectral points at lower wavelengths are included in the patch. We pad l zeros to the start and the end of corrected spectra. This patch with a size of $c = 2l + 1$ serves as the token of the centering spectral point at λ_j. The tokenization process generates a sequence of patches $\boldsymbol{\mathcal{X}}_{(c)} \in \mathbb{R}^{n\times b\times c}$, which is utilized as the spectral embeddings after position encoding.

As mentioned above, baselines $\mathcal{B}$ are likely to contain discriminative information due to the estimation error. The existing baseline corrector could "over-correct" spectra, affecting subsequent steps. We therefore restore baselines after the

corrected spectra are encoded by spectral-attention layer:

$$\begin{aligned}\boldsymbol{\mathcal{X}}_{en} &= \mathrm{SpectrAttn}(\boldsymbol{\mathcal{X}}_{(c)}) \\ \boldsymbol{X}_{en} &= \alpha_1 \boldsymbol{\mathcal{X}}_{en} + \alpha_2 \boldsymbol{\mathcal{B}},\end{aligned} \tag{2}$$

where $\boldsymbol{\mathcal{X}}_{en}$ is the output of spectral-attention layer. $\mathrm{SpectrAttn}$ is the notations of spectral-attention, the details of which will be introduced in the following subsection. Two trainable parameters, namely α_1 and α_2, are introduced to adjust the influence of baselines.

We also introduce a trainable label projecting module for regression tasks, which are ubiquitous in IR spectra analysis (e.g. quantitative estimation of chemical contents). The label projecting module records the statistical properties of training labels, namely mean μ and range r. Given a network output $\hat{\boldsymbol{\mathcal{Y}}}$, the label projecting module maps the output into the original label space:

$$\hat{\boldsymbol{Y}} = (\tanh(\hat{\boldsymbol{\mathcal{Y}}}) - \alpha_3) * r/\alpha_4 + \mu, \tag{3}$$

where $\hat{\boldsymbol{Y}} \in \mathbb{R}^{n\times 1}$ denotes the final prediction. α_3 and α_4 are two trainable parameters that provide flexibility to the projecting module, while $\tanh$ is introduced for scaling and nonlinearity.

Spectral-Attention

The inter-spectral correlations are vital to IR spectroscopy. To enhance the ability to capture such correlations, we introduce inter-spectral attention. We select n_r spectra from the calibration (training) set and form a set of reference spectra $\boldsymbol{X}_r$. The reference spectra are corrected and embedded by the above learnable spectral processing module, deriving reference patches $\boldsymbol{\mathcal{X}}_{r(c)} \in \mathbb{R}^{n_r\times b\times c}$. $\boldsymbol{\mathcal{X}}^{(:,\lambda_j,:)}$ denotes the slice of $\boldsymbol{\mathcal{X}}$ at wavelength λ_j. Given a batch of processed data $\boldsymbol{\mathcal{X}}_{(c)}$, the inter-spectral attention (referred as $\mathrm{InterSpecAttn}$) at wavelength λ_j is computed as:

$$\begin{aligned}&\mathrm{InterSpecAttn}(\boldsymbol{\mathcal{X}}_{(c)}^{(:,\lambda_j,:)}, \boldsymbol{\mathcal{X}}_{r(c)}^{(:,\lambda_j,:)}) \\ &= \mathrm{Softmax}(\mathrm{Filter}(\frac{\boldsymbol{Q}\boldsymbol{K}_{\mathrm{cat}}^T}{\sqrt{d_k}}))\boldsymbol{V}_{\mathrm{cat}} \\ &\boldsymbol{K}_{\mathrm{cat}} = \mathrm{Concat}(\boldsymbol{K}, \boldsymbol{K}_r) \quad \boldsymbol{V}_{\mathrm{cat}} = \mathrm{Concat}(\boldsymbol{V}, \boldsymbol{V}_r),\end{aligned} \tag{4}$$

where $\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V} \in \mathbb{R}^{n\times d_k}$ are the queries, keys and values of $\boldsymbol{\mathcal{X}}_{(c)}^{(:,\lambda_j,:)} \in \mathbb{R}^{n\times c}$ respectively, while $\boldsymbol{K}_r, \boldsymbol{V}_r \in \mathbb{R}^{n_r\times d_k}$ are the keys and values of $\boldsymbol{\mathcal{X}}_{r(c)}^{(:,\lambda_j,:)}$. Operation Filter will suppress the attention between spectra within $\boldsymbol{\mathcal{X}}$ to 0, highlighting the attention from reference spectra. Reference spectra could serve as the "memory" of ACT. The domain-specific information has been eliminated by spectral processing and these corrected spectra could regularize the attention weights to learn domain-invariant features.

Inter-spectral attention calculates the attention between input samples and reference samples. Residual connection is also introduced and the output of inter-spectral attention is given as:

$$\begin{aligned}&\boldsymbol{\mathcal{X}}_{\mathrm{Inter}} = \\ &\mathrm{Concat}(\mathrm{InterSpecAttn}(\boldsymbol{\mathcal{X}}^{(:,\lambda_j,:)}, \boldsymbol{\mathcal{X}}_r^{(:,\lambda_j,:)})_{j=1}^b) + \boldsymbol{\mathcal{X}}.\end{aligned} \tag{5}$$

A single absorbance/reflectance peak covers several adjacent spectral bands, resulting in correlations between these bands. Non-adjacent spectral peaks might also correlate with each other, as a single chemical compound could result in multiple absorbance peaks. We introduce a correlation matrix $\boldsymbol{C} \in \mathbb{R}^{b\times b}$ to record the correlation between spectral bands and utilize it to guide intra-spectral attention. For data $\boldsymbol{\mathcal{X}}$, the intra-spectral attention of i th sample $\boldsymbol{\mathcal{X}}_{(c)}^{(i,:,:)} \in \mathbb{R}^{b\times c}$ is defined as:

$$\begin{aligned}&\mathrm{IntraSpecAttn}(\boldsymbol{\mathcal{X}}_{(c)}^{(i,:,:)}, \boldsymbol{C}) \\ &= \mathrm{Softmax}((1-\alpha_5) * \frac{\boldsymbol{Q}\boldsymbol{K}^T}{\sqrt{d_k}} + \alpha_5 * \frac{\boldsymbol{C}\boldsymbol{W}_c}{\sqrt{b}})\boldsymbol{V},\end{aligned} \tag{6}$$

where $\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V} \in \mathbb{R}^{b\times d_k}$ are the queries, keys and values of $\mathcal{X}_{(c)}^{(i,:,:)}$ respectively. $\boldsymbol{W}_c \in \mathbb{R}^{b\times b}$ is a learnable projection matrix, while α_5 is also a learnable weight that adjusts the influence of correlation matrix. In this paper, $\boldsymbol{\mathcal{X}}_{\mathrm{Inter}}$ serves as the input of intra-spectral attention, and the corresponding representation is given as:

$$\boldsymbol{\mathcal{X}}_{\mathrm{Intra}} = \mathrm{Concat}(\mathrm{IntraSpecAttn}(\boldsymbol{\mathcal{X}}_{\mathrm{Inter}}^{(i,:,:)}, \boldsymbol{C})_{i=1}^n). \tag{7}$$

$\mathrm{IntraSpecAttn}$ calculates 'global' attention across the whole spectrum, which could capture the long-range dependencies among spectral peaks.

The final output of spectral-attention is the combination of inter- and intra-spectral attention, where another learnable parameter is introduced to control the influences of the two attentions. A fully connected feed-forward network with batch normalization is also included:

$$\boldsymbol{\mathcal{X}}_{\mathrm{SpectrA}} = \mathrm{FeedForward}((1-\alpha_6)\boldsymbol{\mathcal{X}}_{\mathrm{Inter}} + \alpha_6 * \boldsymbol{\mathcal{X}}_{\mathrm{Intra}}), \tag{8}$$

where α_6 is a learnable weight that adjusts influences of inter- and intra- spectral attention. It should be noted that we use batch normalization instead of layer normalization between the feed-forward network to increase the interactions between spectra. The encoded feature $\boldsymbol{\mathcal{X}}_{\mathrm{SpectrA}} \in \mathbb{R}^{n\times b\times c}$ is then compressed into $\boldsymbol{\mathcal{X}}_{en} \in \mathbb{R}^{n\times b}$ with a weighted sum process along the token dimension. Two learnable projection matrix $\boldsymbol{W}_v, \boldsymbol{W}_w \in \mathbb{R}^{c\times c}$ are introduced:

$$\begin{aligned}&\boldsymbol{\mathcal{X}}_{en} = \sum_{k=1}^{c} \mathrm{Compress}(\boldsymbol{\mathcal{X}}_{\mathrm{SpectrA}})^{(:,:,k)}, \\ &\mathrm{Compress}(\boldsymbol{\mathcal{X}}_{\mathrm{SpectrA}}) = \\ &\mathrm{Concat}((\boldsymbol{\mathcal{X}}_{\mathrm{SpectrA}}^{(i,:,:)}\boldsymbol{W}_v \odot \mathrm{Softmax}(\boldsymbol{\mathcal{X}}_{\mathrm{SpectrA}}^{(i,:,:)}\boldsymbol{W}_w))_{i=1}^n),\end{aligned} \tag{9}$$

where $\odot$ denotes Hadamard Product or element-wise product. $\mathrm{Compress}(\boldsymbol{\mathcal{X}}_{\mathrm{SpectrA}}) \in \mathbb{R}^{n\times b\times c}$ represents the weighted $\boldsymbol{\mathcal{X}}_{\mathrm{SpectrA}}$ where the significance of each token features has been added. The above compression operator calculates correlations within sliding windows. The encoded feature $\boldsymbol{\mathcal{X}}_{en}$ is then handled by the learnable spectral processing module and fully connected layers to get the final prediction of ACT.

The complexity of spectral-attention is $O(b^3)$ in theory, and the number of spectral bands b is limited due to the constraints of optical resolution. The extra computation consumption of spectral-attention is therefore limited.

	Train: Val: Test	Type
Tablet(1,2)	155: 40: 460	Regression(CT)
Tablet(2,1)	155: 40: 460	Regression(CT)
MF(R562,R568)	2122: 910: 733	Regression(CT)
MF(R568,R562)	513: 220: 3032	Regression(CT)
Tablet(1,1)	155: 40: 460	Regression
Tablet(2,2)	155: 40: 460	Regression
Mango_DMC	7413: 2830: 1448	Regression
Strawberry	337: 329: 317	Classification
Apple_leaf	2500: 1250: 1740	Classification

Table 1: Datasets and corresponding setups. CT stands for *calibration transfer* tasks, while the others are *regular* spectra modeling tasks.

Experiment

Experimental Setup

Datasets The 9 real-world tasks analyzed in this study are derived from 5 open-source datasets. We utilize the standard training, validation, and testing sets provided by these datasets. (1) Tablet dataset (Nikzad-Langerodi et al. 2020) contains NIR spectra of tablets collected by two individual spectrometers (referred to as spectrometer No. 1 and No. 2). Tablet dataset gives 4 active pharmaceutical ingredient (API) prediction tasks: Tablet(1, 1), Tablet(1, 2), Tablet(2, 2) and Tablet(2, 1). Tablet(1, 2) means that the calibrating (training and validating) data is from spectrometer No. 1 and the testing data is from spectrometer No. 2. (2) Melamine dataset (Nikzad-Langerodi et al. 2018) consists of NIR spectra collected from melamine-formaldehyde with slightly different compositions. We use two recipes R562 and R568 to generate 2 tasks for turbidity point prediction: MF(R562, R568) and MF(R568, R562). Notably, the two MF tasks include selected characteristic bands rather than whole spectra. (3) Mango_DMC dataset (Anderson et al. 2020) contains NIR spectra of intact mango fruit, aiming at predicting the dry matter content across different seasons, locations, and cultivars. (4) Strawberry dataset (Holland, Kemsley, and Wilson 1998) tries to classify the MIR spectra of strawberry purees from those of other fruit purees, with a FTIR spectrometer. (5) Apple_Leaf (Xue et al. 2016) aims to classify NIR spectra of apple leaves from 20 different varieties or cultivars. The datasets are summarized in Table 1.

Baselines We include 4 deep networks and 4 traditional calibration transfer methods for comparison. (1) DeepSpectra (Zhang et al. 2019) is a far-reaching end-to-end network for quantitative spectral analysis, which is based on the Inception network. (2) AggMapNet (Wang et al. 2023) converts infrared spectra into 2D maps for feature augmentation and introduces 2D CNN for learning the maps. (3) TeaNet (Wu et al. 2024) masks and reconstructs the input spectra for data augmentation, where the augmented data are used to boost the modeling performance. (4) Spectraformer (Chen, Zhou, and Ren 2024) is a hybrid network for IR spectroscopy that combines 1D convolutional layers with an attention layer. (5) PLS (Mishra et al. 2022c), namely partial least squares, is widely used for regression tasks of IR analysis. (6) di-PLS (domain invariant PLS) (Nikzad-Langerodi et al. 2018) introduces a domain regularizer for calibration transfer. (7) JDOT (Courty et al. 2017) can handle shifts in both margin and conditional distributions. (8) DIPALS (Nikzad-Langerodi et al. 2020) is a domain adaptation method for IR spectra.

	l	n_h	d_{ff}	FC_layers	n_r
Tablet	12	5	512	[256,64]	36
Mango_DMC	12	5	512	[256,64]	36
Strawberry	12	5	512	[256,64]	36
Melamine	2	1	128	[64, 16]	36
Apple_leaf	22	1	256	[128, 32]	36

Table 2: Hyper-parameter settings of ACT. d_{ff} stands for the dimension of feed forward network, while n_h is the number of attention heads.

Implementation Details Hyper-parameter settings of ACT is listed in Table 2. The datasets vary in spectral range, resolution, etc., necessitating the tuning of hyper-parameters for each dataset to prevent significant over-fitting or under-fitting. In all the tasks, the hyper-parameters of tested methods are adjusted based on data in the calibrating sets. Notably, when dealing with tasks from the same dataset, the hyper-parameters of deep networks remain unchanged. We use root mean square error of prediction (RMSEP) and mean absolute error (MAE) as the evaluation metrics for regression, while accuracy (ACC) and weighted F1 (F1) score are used for classification. All the experiments are implemented based on PyTorch (Paszke et al. 2019) and are repeated 5 times with NVIDIA RTX 4090 24GB GPU.

Calibration Transfer

We report the results of calibration transferring tasks in Table 3 and Table 4. The tested DL methods are trained without any access to the spectra in the target domain (domain generalization), while the traditional calibration methods utilize 60% unlabeled data from the target domain.

As shown in Table 3, it is indicated that the proposed ACT outperforms DL methods across various tasks with a considerable margin in terms of both RMSEP and MAE. Specifically, ACT achieves a 27% reduction of RMSEP and a 23% reduction of MAE on Tablet tasks. Compared with the classical calibration transfer methods, ACT achieves comparable (even better) results without access to target domain data, as presented in Table 4. Results from (Nikzad-Langerodi et al. 2020) are used in this table. It is indicated that ACT could learn domain-invariant representations without access to the target domain. Eliminating the need for target domain samples means lower cost and higher efficiency. Moreover, the time cost in Table 5 also proves that the efficiency of ACT is similar to that of comparison networks.

ACT might owe this improvement to the combination of Transformer and knowledge of analytical chemistry. Transformer provides abundant learning ability beyond classical methods, ensuring model robustness across different scenarios. On Tablet tasks, Spectraformer, which includes the en-

	Tablet(1, 2)		Tablet(2, 1)		MF(R562,R568)		MF(R568,R562)	
	RMSEP	MAE	RMSEP	MAE	RMSEP	MAE	RMSEP	MAE
DeepSpectra	11.676 ±3.275	9.762 ±3.013	14.983 ±5.08	13.729 ±5.272	4.055 ±1.148	3.239 ±1.043	4.507 ±1.736	3.763 ±1.713
AggMapNet	10.142 ±0.901	7.031 ±0.746	13.212 ±1.154	10.322 ±1.111	2.631 ±0.227	**1.697** ±**0.131**	3.775 ±0.124	2.827 ±0.125
TeaNet	11.819 ±3.673	9.758 ±3.654	17.944 ±4.112	14.961 ±4.159	5.733 ±1.586	4.424 ±1.414	10.301 ±2.374	8.469 ±2.053
Spectraformer	8.810 ±2.048	6.497 ±1.837	9.279 ±2.239	7.445 ±2.272	3.168 ±0.633	2.563 ±0.561	4.247 ±0.451	3.384 ±0.254
ACT	**6.941** ±**3.255**	**5.210** ±**2.994**	**6.155** ±**0.985**	**4.786** ±**1.156**	**2.293** ±**0.624**	1.798 ±0.51	**2.223** ±**0.202**	**1.614** ±**0.199**
Imp(%)	21.2%	19.8%	33.7%	35.7%	12.8%	-6.0%	41.1%	42.9%

Table 3: Comparison with deep methods on *calibration transfer* tasks without any access to the target domain. IMP(%) stands for relative improvements and the best results are highlighted in *bold*.

	Tablet(1,2)	Tablet(2,1)	MF(R562,R568)	MF(R562,R568)
di-PLS	8.690 ± 1.060	7.980 ± 0.830	2.470 ± 0.984	2.580 ± 0.997
JDOT	12.860 ± 1.070	13.460 ± 0.500	3.230 ± 0.300	2.950 ± 0.150
DIPALS	7.690 ± 0.470	7.120 ± 0.680	1.750 ± 0.140	2.202 ± 0.140
PLS	9.038 ± 0.000	16.702 ± 0.000	2.690 ± 0.000	2.630 ± 0.000
ACT	6.941 ± 3.255	5.210 ± 2.994	2.293 ± 0.624	2.223 ± 0.202

Table 4: Comparison with classical *calibration transfer* methods in terms of RMSEP. Notably, methods marked in *italic* achieve the results *with* 60% of unlabeled target domain data, while PLS and ACT are trained *without* the target domain data. As standard data segmentation is used in all the experiments, the performance of PLS is invariant in the repeated experiments.

	Tablet(1,2)		MF(R562,R568)	
	Train(s)	Test(s)	Train(s)	Test(s)
DeepSpectra	0.0108	0.0140	0.2201	0.0219
AggMapNet	0.0798	0.0630	0.2500	0.0817
TeaNet	0.1850	0.2193	1.4657	0.3189
Spectraformer	0.0180	0.0359	0.2264	0.0817
ACT	0.0960	0.1114	0.4369	0.0698

Table 5: Time cost (in seconds) per epoch of deep networks.

coder of Vanilla Transformer, achieves results comparable to those of traditional calibration transfer methods. The prior knowledge further regularizes ACT to learn representations in accord with the intrinsic meaning of spectra (proved in the following visualization results). Specifically, learnable spectra processing could alleviate the effect of 'over-correct', while spectral-attention introduces additional inductive bias.

Regular IR Spectral Data Analysis

The experimental results on regular IR spectral data analysis are reported in Table 6. The proposed ACT is further evaluated on regular IR spectral datasets where spectra are collected in relatively stationary environments. In both qualitative (classification) and quantitative (regression) tasks, ACT achieves competitive results across different datasets. Specifically, ACT achieves an average rank of 1.2 on both the regression and the classification tasks, while the second-best average rank is 3.1. It is indicated that incorporating prior knowledge can also improve the performance on regular datasets, as the representations in accord with the principles of IR usually have better generalization ability.

Ablation Results

The ablation study is conducted on: Tablet(1, 2) for calibration transfer, Mango_DMC for regular regression, and Strawberry for regular classification. Several hierarchical models are used: (1) *Base* is the basic model containing an encoder layer based on self-attention and a fully connected network with 2 hidden layers. (2) *Base + Token* is integrated with the proposed tokenization method. (3) *Base + Token + LearnProc'* further incorporates the learnable spectral processing module without baseline reconstruction (Eq. 2). (4) *Base + Token + LearnProc* introduce original learnable spectral processing. (5) *Base + Token + SpectrAttn* is the combination of base model and spectral-attention. (6) *IntraSpec, InterSpec* evaluate the impact of individual intra- and inter-spectral attention. (7) *SNV* and *Derivative* (Mishra et al. 2022c), two traditional pre-processing methods, are used to evaluate the effectiveness of *LearnProc*.

Ablation results in Table 7 indicate the effectiveness of proposed modules. The learnable spectral processing enables ACT to reclaim useful information from the removed baselines. Moreover, it also enhances the adaptability of pre-processing and post-processing. It is indicated that both intra- and inter-spectral attention are effective. The two components can improve performance in certain scenarios but lack adaptability across different tasks. Spectral-attention,

	Tablet(1, 1)		Tablet(2, 2)		Mango_DMC		Strawberry_puree		Apple_leaf		
	RMSEP	MAE	RMSEP	MAE	RMSEP	MAE	ACC	F1	ACC	F1	Rank
DeepSpectra	6.218	4.269	5.738	4.054	1.039	0.791	0.960	0.960	0.661	0.608	3.8
	±1.034	±0.941	±0.676	±0.475	±0.09	±0.048	±0.008	±0.007	±0.051	±0.061	
AggMapNet	5.543	4.007	6.329	4.238	1.357	0.97	0.962	0.962	0.424	0.405	3.1
	±0.493	±0.609	±0.379	±0.323	±0.047	±0.018	±0.002	±0.002	±0.004	±0.006	
TeaNet	5.972	4.093	6.194	4.289	1.143	0.892	0.921	0.921	0.785	**0.791**	3.5
	±0.548	±0.392	±0.154	±0.166	±0.151	±0.067	±0.011	±0.011	±0.046	**±0.045**	
Spectraformer	6.679	4.404	5.956	3.958	1.038	0.781	0.960	0.959	0.502	0.450	3.4
	±0.839	±0.466	±0.242	±0.191	±0.067	±0.046	±0.009	±0.009	±0.038	±0.046	
ACT	**4.604**	**2.914**	**4.417**	**2.714**	**1.008**	**0.756**	**0.965**	**0.965**	**0.793**	0.779	**1.2**
	±0.661	**±0.688**	**±0.150**	**±0.186**	**±0.057**	**±0.039**	**±0.004**	**±0.004**	**±0.014**	±0.013	

Table 6: *Regular* IR spectral data analysis. Rank denotes the average rank and the best results are highlighted in bold.

	Tablet(1, 2)		Mango_DMC		Strawberry_puree		Rank
	RMSEP	MAE	RMSEP	MAE	ACC	F1	
Base	9.977	7.521	1.706	1.277	0.907	0.904	10.00
Base + Token	8.491	6.477	1.130	0.854	0.953	0.953	7.33
Base + Token + LearnProc'	8.054	5.611	1.459	1.099	0.953	0.953	7.33
Base + Token + LearnProc	7.833	5.690	1.083	0.813	0.951	0.951	6.50
Base + Token + IntraSpec	7.562	4.874	1.348	0.973	0.962	0.962	5.50
Base + Token + InterSpec	7.265	5.389	1.151	0.858	0.969	0.969	3.83
Base + Token + SpectrAttn	7.152	4.941	1.009	0.747	0.962	0.962	2.83
Base + Token + SpectrAttn + SNV	9.077	6.962	1.055	0.774	0.965	0.965	5.33
Base + Token + SpectrAttn + Derivative	8.032	6.069	1.055	0.774	0.968	0.968	4.00
ACT (Base + Token + SpectrAttn + LearnProc)	6.941	5.210	1.008	0.756	0.965	0.965	2.33

Table 7: Ablation study. Notation '+' means incorporating specific modules.

the combination of intra- and inter-attention, has shown more stable performance. Meanwhile, the results also prove the effectiveness of baseline restoration (Eq. 2) and the superiority of learnable spectral processing over the traditional pre-processing methods.

The effect of random seed for reference selection is also evaluated on the Tablet(1,2) task, and RMSEP varies from 5.59 to 7.78. The effect of baseline correction could vary and some processed spectra may therefore contain more domain-invariant information than others. ACT therefore would perform better when using those spectra as "memory" and vice versa. When target domain data are directly used for training, ACT performs obviously better on Tablet(1,2) in terms of RMSEP (6.94 to 5.10) and MAE (5.21 to 3.46).

Visualization

To visualize ACT, attention maps, along with original spectra, of Tablet(1,2) and Mango_DMC dataset are presented in Fig. 3. It is indicated that ACT only generates higher activation at some of the peaks rather than all the peaks, which is similar to the analysis results of experts. On Mango_DMC, there is high activation around the absorbance peak of chlorophyll (680 nm) and O-H (800 – 1000 nm). The distribution of attention is generally in accord with the key absorbance peaks identified by human experts (Anderson et al. 2020). Moreover, there are higher attention values at the wavelengths where spectra have large variance. In this sense, spectral-attention seems to be able to learn representations that are both domain-invariant and aligned with the experts' judgement.

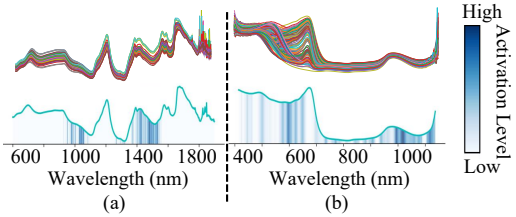

Figure 3: Original spectra (upper) and corresponding attention map (lower) on: (a) Tablet(1, 2) and (b) Mango_DMC.

Conclusion

This paper studies the integration of deep learning and analytical chemistry knowledge. The calibration transfer problem and the distinctive properties of IR spectra hinder the further application of deep neural networks in IR spectroscopy. We propose the ACT, a deep learning method that incorporates prior knowledge of IR spectroscopy. For ACT, we designed two modules informed by analytical chemistry, namely learnable spectral processing and spectral-attention. ACT is evaluated in 9 tasks and achieves competitive performance, exhibiting the ability to learn domain-invariant representations. We believe this work promotes the development of deep learning in IR spectroscopy.

Acknowledgments

This research is supported by the Science and Technology Plan of Huzhou Science and 319 Technology Bureau (No. 2021GZB01).

References

Anderson, N.; Walsh, K.; Subedi, P.; and Hayes, C. 2020. Achieving robustness across season, location and cultivar for a NIRS model for intact mango fruit dry matter content. *Postharvest Biology and Technology*, 168: 111202.

Bredenbeck, J.; Ghosh, A.; Smits, M.; and Bonn, M. 2008. Ultrafast Two Dimensional-Infrared Spectroscopy of a Molecular Monolayer. *Journal of the American Chemical Society (JACS)*, 130(7): 2152–2153.

Chen, Y.-Y.; and Wang, Z.-B. 2019. End-to-end quantitative analysis modeling of near-infrared spectroscopy based on convolutional neural network. *Journal of Chemometrics*, 33(5): e3122.

Chen, Z.; Zhou, R.; and Ren, P. 2024. Spectraformer: deep learning model for grain spectral qualitative analysis based on transformer structure. *RSC Advances*, 14: 8053–8066.

Courty, N.; Flamary, R.; Habrard, A.; and Rakotomamonjy, A. 2017. Joint distribution optimal transportation for domain adaptation. In Guyon, I.; Luxburg, U. V.; Bengio, S.; Wallach, H.; Fergus, R.; Vishwanathan, S.; and Garnett, R., eds., *31st Annual Conference on Neural Information Processing Systems (NIPS)*, volume 30. Curran Associates, Inc.

Cui, C.; and Fearn, T. 2018. Modern practical convolutional neural networks for multivariate regression: Applications to NIR calibration. *Chemometrics and Intelligent Laboratory Systems*, 182: 9–20.

Debus, B.; Parastar, H.; Harrington, P.; and Kirsanov, D. 2021. Deep learning in analytical chemistry. *TrAC Trends in Analytical Chemistry*, 145: 116459.

Enders, A. A.; North, N. M.; Fensore, C. M.; Velez-Alvarez, J.; and Allen, H. C. 2021. Functional Group Identification for FTIR Spectra Using Image-Based Machine Learning Models. *Analytical Chemistry*, 93(28): 9711–9718. PMID: 34190551.

Gan, F.; Ruan, G.; and Mo, J. 2006. Baseline correction by improved iterative polynomial fitting with automatic threshold. *Chemometrics and Intelligent Laboratory Systems*, 82(1): 59–65. Selected Papers from the International Conference on Chemometrics and Bioinformatics in Asia.

Holland, J. K.; Kemsley, E. K.; and Wilson, R. H. 1998. Use of Fourier transform infrared spectroscopy and partial least squares regression for the detection of adulteration of strawberry purées. *Journal of the Science of Food and Agriculture*, 76(2): 263–269.

John-Herpin, A.; Tittl, A.; Kühner, L.; Richter, F.; Huang, S. H.; Shvets, G.; Oh, S.-H.; and Altug, H. 2023. Metasurface-Enhanced Infrared Spectroscopy: An Abundance of Materials and Functionalities. *Advanced Materials*, 35(34): 2110163.

Jung, G.; Jung, S. G.; and Cole, J. M. 2023. Automatic materials characterization from infrared spectra using convolutional neural networks. *Chemical Secience*, 14: 3600–3609.

Kalashnikov, D. A.; Paterova, A. V.; Kulik, S. P.; and Krivitsky, L. A. 2016. Infrared spectroscopy with visible light. *Nature Photonics*, 10(2): 98–101.

LeCun, Y.; Bengio, Y.; and Hinton, G. 2015. Deep learning. *Nature*, 521(7553): 436–444.

Li, L.; Zang, H.; Li, J.; Chen, D.; Li, T.; and Wang, F. 2014. Identification of anisodamine tablets by Raman and near-infrared spectroscopy with chemometrics. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 127: 91–97.

Liu, Z.; Cheng, M.; Li, Z.; Huang, Z.; Liu, Q.; Xie, Y.; and Chen, E. 2024. Adaptive normalization for non-stationary time series forecasting: A temporal slice perspective. *In NeurIPS 2023*, 36.

Manley, M. 2014. Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials. *Chemical Society Reviews*, 43: 8200–8214.

Marini, F. 2009. Artificial neural networks in foodstuff analyses: Trends and perspectives A review. *Analytica Chimica Acta*, 635(2): 121–131.

Martens, J.; Koppen, V.; Berden, G.; Cuyckens, F.; and Oomens, J. 2017. Combined Liquid Chromatography-Infrared Ion Spectroscopy for Identification of Regioisomeric Drug Metabolites. *Analytical Chemistry*, 89(8): 4359–4362.

Miao, T.; Sihota, N.; Pfeifer, F.; McDaniel, C.; De Gea Neves, M.; and Siesler, H. W. 2023. Rapid Determination of the Total Petroleum Hydrocarbon Content of Soils by Handheld Fourier Transform Near-Infrared Spectroscopy. *Analytical Chemistry*, 95(17): 6888–6893. PMID: 37070825.

Mishra, P.; Nikzad-Langerodi, R.; Marini, F.; Roger, J. M.; Biancolillo, A.; Rutledge, D. N.; and Lohumi, S. 2021. Are standard sample measurements still needed to transfer multivariate calibration models between near-infrared spectrometers? The answer is not always. *TrAC Trends in Analytical Chemistry*, 143: 116331.

Mishra, P.; and Passos, D. 2021. Deep calibration transfer: transferring deep learning models between infrared spectroscopy instruments. *Infrared Physics & Technology*, 117: 103863.

Mishra, P.; Passos, D.; Marini, F.; Xu, J.; Amigo, J. M.; Gowen, A. A.; Jansen, J. J.; Biancolillo, A.; Roger, J. M.; Rutledge, D. N.; and Nordon, A. 2022a. Deep learning for near-infrared spectral data modelling: Hypes and benefits. *TrAC Trends in Analytical Chemistry*, 157: 116804.

Mishra, P.; Passos, D.; Marini, F.; Xu, J.; Amigo, J. M.; Gowen, A. A.; Jansen, J. J.; Biancolillo, A.; Roger, J. M.; Rutledge, D. N.; and Nordon, A. 2022b. Deep learning for near-infrared spectral data modelling: Hypes and benefits. *TrAC Trends in Analytical Chemistry*, 157: 116804.

Mishra, P.; Roger, J. M.; Marini, F.; Biancolillo, A.; and Rutledge, D. N. 2022c. Pre-processing ensembles with response oriented sequential alternation calibration (PROSAC): A step towards ending the pre-processing search and optimization quest for near-infrared spectral modelling. *Chemometrics and Intelligent Laboratory Systems*, 222: 104497.

Mokari, A.; Guo, S.; and Bocklitz, T. 2023. Exploring the Steps of Infrared (IR) Spectral Analysis: Pre-Processing, (Classical) Data Modelling, and Deep Learning. *Molecules*, 28(19).

Nie, Y.; Nguyen, N. H.; Sinthong, P.; and Kalagnanam, J. 2022. A time series is worth 64 words: Long-term forecasting with transformers. *In ICLR 2023*.

Nikzad-Langerodi, R.; Zellinger, W.; Lughofer, E.; and Saminger-Platz, S. 2018. Domain-Invariant Partial-Least-Squares Regression. *Analytical Chemistry*, 90(11): 6693–6701.

Nikzad-Langerodi, R.; Zellinger, W.; Saminger-Platz, S.; and Moser, B. A. 2020. Domain adaptation for regression under Beer–Lambert's law. *Knowledge-Based Systems*, 210: 106447.

Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *In NeurIPS 2019*, 32.

Peng, J.; Peng, S.; Xie, Q.; and Wei, J. 2011. Baseline correction combined partial least squares algorithm and its application in on-line Fourier transform infrared quantitative analysis. *Analytica chimica acta*, 690(2): 162–168.

Porep, J. U.; Kammerer, D. R.; and Carle, R. 2015. On-line application of near infrared (NIR) spectroscopy in food production. *Trends in Food Science & Technology*, 46(2, Part A): 211–230.

Ramzan, M.; Raza, A.; un Nisa, Z.; Abdel-Massih, R. M.; Al Bakain, R.; Cabrerizo, F. M.; Dela Cruz, T. E.; Aziz, R. K.; and Musharraf, S. G. 2024. Detection of antimicrobial resistance (AMR) and antimicrobial susceptibility testing (AST) using advanced spectroscopic techniques: A review. *TrAC Trends in Analytical Chemistry*, 172: 117562.

Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; and Wojna, Z. 2016. Rethinking the inception architecture for computer vision. In *In CVPR 2016*, 2818–2826.

Varmuza, K.; and Filzmoser, P. 2016. *Introduction to multivariate statistical analysis in chemometrics*. CRC press.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *In NeurIPS 2017*, 30.

Vettoruzzo, A.; Bouguelia, M.-R.; Vanschoren, J.; Rögnvaldsson, T.; and Santosh, K. 2024. Advances and Challenges in Meta-Learning: A Technical Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 46(7): 4763–4779.

Wang, T.; Tan, Y.; Chen, Y. Z.; and Tan, C. 2023. Infrared Spectral Analysis for Prediction of Functional Groups Based on Feature-Aggregated Deep Learning. *Journal of Chemical Information and Modeling*, 63(15): 4615–4622.

Wang, X.; Jiang, S.; Hu, W.; Ye, S.; Wang, T.; Wu, F.; Yang, L.; Li, X.; Zhang, G.; Chen, X.; Jiang, J.; and Luo, Y. 2022a. Quantitatively Determining Surface–Adsorbate Properties from Vibrational Spectroscopy with Interpretable Machine Learning. *Journal of the American Chemical Society (JACS)*, 144(35): 16069–16076.

Wang, Z.; Zhang, J.; Zhang, X.; Chen, P.; and Wang, B. 2022b. Transformer model for functional near-infrared spectroscopy classification. *IEEE Journal of Biomedical and Health Informatics*, 26(6): 2559–2569.

Weber, A.; Hoplight, B.; Ogilvie, R.; Muro, C.; Khandasammy, S. R.; Pérez-Almodóvar, L.; Sears, S.; and Lednev, I. K. 2023. Innovative Vibrational Spectroscopy Research for Forensic Application. *Analytical Chemistry*, 95(1): 167–205. PMID: 36625116.

Wu, H.; Xu, J.; Wang, J.; and Long, M. 2021. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *In NeurIPS 2021*, 34: 22419–22430.

Wu, Y.; Liu, J.; Wang, Y.; Gibson, S.; Osadchy, M.; and Fang, Y. 2024. Reconstructing Randomly Masked Spectra Helps DNNs Identify Discriminant Wavenumbers. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 46(5): 3845–3861.

Xue, X.; Zhuang, L.; Dingfeng, W.; and Xiaoli, W. 2016. Image and Spectral Dataset of Standard Type Blades of Apple Varieties.

Yang, X.; Zhuang, X.; Shen, R.; Sang, M.; Meng, Z.; Cao, G.; Zang, H.; and Nie, L. 2024. In situ rapid evaluation method of quality of peach kernels based on near infrared spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 313: 124108.

Yu, H.; Du, W.; Lang, Z.-Q.; Wang, K.; and Long, J. 2021a. A Novel Integrated Approach to Characterization of Petroleum Naphtha Properties From Near-Infrared Spectroscopy. *IEEE Transactions on Instrumentation and Measurement*, 70: 1–13.

Yu, Y.; Huang, J.; Liu, S.; Zhu, J.; and Liang, S. 2021b. Cross target attributes and sample types quantitative analysis modeling of near-infrared spectroscopy based on instance transfer learning. *Measurement*, 177: 109340.

Zhang, J.; Zhou, X.; and Li, B. 2023. PFCE2: A versatile parameter-free calibration enhancement framework for near-infrared spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 301: 122978.

Zhang, X.; Lin, T.; Xu, J.; Luo, X.; and Ying, Y. 2019. DeepSpectra: An end-to-end deep learning approach for quantitative spectral analysis. *Analytica Chimica Acta*, 1058: 48–57.

Åsmund Rinnan; van den Berg, F.; and Engelsen, S. B. 2009. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry*, 28(10): 1201–1222.