



Integrating transformer-based machine learning with SERS technology for the analysis of hazardous pesticides in spinach

Mehdi Hajikhani^{a,1}, Akashata Hegde^{b,1}, John Snyder^c, Jianlin Cheng^{b,d,*}, Mengshi Lin^{a,**}

^a Food Science Program, University of Missouri, Columbia, MO 65211, USA

^b Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO 65211, USA

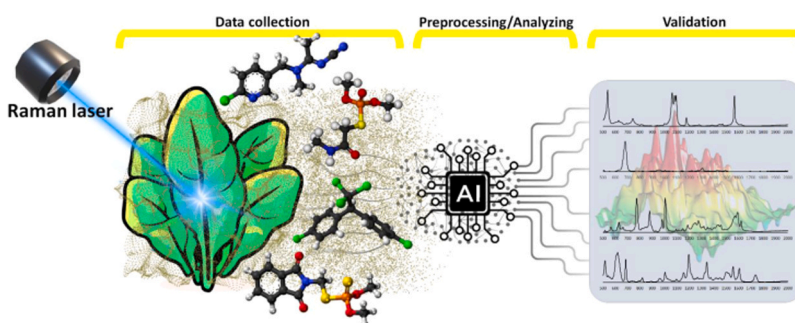
^c Department of Statistics, University of Missouri, Columbia, MO 65211, USA

^d Roy Blunt Next Gen Precision Health, University of Missouri, Columbia, MO 65201, USA

HIGHLIGHTS

- A novel SERS method coupled with machine learning tools was established.
- Gold-silver core-shell nanoparticles significantly enhanced Raman signals.
- Five types of pesticides were classified by mL models with 98% accuracy.
- Quantitative analysis resulted in an MAE of 0.966 and an MSE of 1.826.
- This study helps monitor pesticide contamination in agricultural products.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:

Pesticide
Artificial intelligence
SERS
Machine learning
Transformer

ABSTRACT

This study introduces an innovative strategy for the rapid and accurate identification of pesticide residues in agricultural products by combining surface-enhanced Raman spectroscopy (SERS) with a state-of-the-art transformer model, termed SERSFormer. Gold-silver core-shell nanoparticles were synthesized and served as high-performance SERS substrates, which possess well-defined structures, uniform dispersion, and a core-shell composition with an average diameter of 21.44 ± 4.02 nm, as characterized by TEM-EDS. SERSFormer employs sophisticated, task-specific data processing techniques and CNN embedders, powered by an architecture features weight-shared multi-head self-attention transformer encoder layers. The SERSFormer model demonstrated exceptional proficiency in qualitative analysis, successfully classifying six categories, including five pesticides (coumaphos, oxamyl, carbophenothion, thiaabendazole, and phosmet) and a control group of spinach data, with 98.4% accuracy. For quantitative analysis, the model accurately predicted pesticide concentrations with a mean absolute error of 0.966, a mean squared error of 1.826, and an R^2 score of 0.849. This novel approach, which combines SERS with machine learning and is supported by robust transformer models, showcases the potential for real-time pesticide detection to improve food safety in the agricultural and food industries.

* Corresponding author at: Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO 65211, USA.

** Corresponding author.

E-mail addresses: chengji@missouri.edu (J. Cheng), linme@missouri.edu (M. Lin).

¹ Co-first authors and equal contributors.

1. Introduction

The widespread use of pesticides in agriculture around the world has been pivotal for enhancing crop yields and ensuring food security [1,2]. However, the persistence of these chemicals raises concerns about residues in food products, posing health risks from acute toxicity to chronic diseases [3,4]. Therefore, there is an urgent need for advanced technologies to quickly and accurately detect pesticide residues, ensuring the safety of agricultural products. In recent years, surface-enhanced Raman spectroscopy (SERS) has emerged as a cutting-edge technique with immense potential to revolutionize food safety assessment. SERS leverages the phenomenon of enhanced Raman scattering when molecules are adsorbed onto metallic nanostructures, such as silver or gold nanoparticles [5]. This method offers many advantages in analyzing pesticide residues in agricultural products [6].

Compared to traditional methods, such as gas chromatography-mass spectrometry (GC-MS) [7], SERS provides non-destructive, rapid analysis, ideal for real-time food industry monitoring [8]. SERS provides high sensitivity for detecting trace chemical compounds with exceptional precision. It reliably identifies different pesticides at low concentrations through the vibrational fingerprints of molecules [9]. Additionally, SERS can be integrated with advanced artificial intelligence (AI) and machine learning techniques based on the cutting-edge transformer technology that has achieved great success in large language models (e.g., ChatGPT) for natural language processing, providing a powerful synergy for accurate and efficient identification of pesticide residues [10,11].

Machine learning, a key subfield of artificial intelligence (AI), develops algorithms that automatically learn hidden patterns and relationships from data without explicit programming [12]. Leveraging big data and high-performance computing, it has revolutionized information processing and decision-making by enabling computers to improve tasks over time [13,14]. Machine learning has significantly evolved from its early focus on pattern recognition to widespread use in various sectors, including healthcare, finance, and food science [15]. In food science, it assists in evaluating dietary consumption across populations [16], improving food processing procedures and safety analysis [17]. Furthermore, deep learning, a key advancement in machine learning featuring multi-layered neural networks, has enhanced its ability to solve complex problems by analyzing intricate patterns in datasets [18]. This progress is largely driven by the availability of large datasets, including those obtained from techniques like SERS.

However, there are challenges in analyzing spectral data, including data preprocessing and variability in Raman signals [19], necessitating robust, specialized models. Additionally, models capable of simultaneously assessing both the quantity and type of pesticide residues in food are crucial. This study investigates combining SERS technology with machine learning, particularly deep learning, for pesticide analysis in foods. This was achieved by developing the SERSFormer that utilizes the innovative transformer deep-learning approach, known for its self-attention mechanism [11], to simultaneously quantify and qualify pesticides in SERS data through multi-task learning and weight sharing [20]. This innovative methodology offers a robust and efficient tool for unraveling complex molecular information encoded in SERS spectra. By harnessing the enhanced sensitivity and specificity of SERS, coupled with the data-processing capabilities of machine learning algorithms, this study aims to pioneer a novel methodology for the rapid and accurate detection of pesticide residues in agricultural products.

2. Experimental

2.1. Materials

In this study, gold(III) chloride solution, silver nitrate, L-ascorbic acid, tri-sodium citrate dihydrate, and 96% ethanol were acquired from Sigma Aldrich (St. Louis, Missouri, USA) to synthesize SERS metallic

nanostructures. Pesticide standards, including coumaphos, oxamyl, carbophenothion, thiabendazole, and phosmet, were procured from Sigma Aldrich in analytical standard grade (PESTANAL®). High-purity chemicals and USDA-approved organic food samples were used to minimize pre-existing residues, focusing the analysis on contaminants introduced rather than those already present. This careful selection ensures a thorough investigation into rapid pesticide residue detection in agricultural products.

2.2. Synthesis of SERS substrates and the characterization of core-shell nanoparticles

The synthesis of gold-silver core-shell nanoparticles (Au@Ag NPs) was carried out to optimize SERS substrate performance [21]. Initially, gold nanoparticles were synthesized by adding 8.6 μL of HAuCl_4 to 50 mL of deionized water. The solution was then heated using a heater stirrer until it approached near-boiling temperature. Subsequently, 1 mL of 1% tri-sodium citrate dihydrate solution was introduced, and the solution was maintained at the boiling temperature for 20 min, resulting in a distinctive color change to wine red. The solution was then removed from heat and allowed to cool. Next, a 100 mM L-ascorbic acid solution was prepared and mixed with the gold nanoparticle solution at a ratio of 1:6, ensuring thorough homogenization for 10 min. Finally, 1 mM solution of AgNO_3 was added drop by drop to the gold solution while stirring at high speed, maintaining a ratio of 1:3.5. The resultant core-shell nanoparticles, crucial for SERS analysis, were obtained after the addition of the silver solution [22]. The core-shell nanoparticles were characterized using transmission electron microscopy with energy-dispersive X-ray spectroscopy (TEM-EDS) using the Spectra 300 STEM instrument (ThermoScientific, Ltd., USA). TEM-EDS analysis reveals the morphology, size, and elemental composition of the gold-silver core-shell nanoparticles, confirming the successful synthesis and validating the structural integrity of the core-shell nanoparticles [22].

2.3. The preparation of food samples and spiking with pesticides

Spinach samples were thoroughly washed, drained, and finely chopped. Pesticide samples, including coumaphos, oxamyl, carbophenothion, thiabendazole, and phosmet, were prepared in concentrations ranging from 0.5 to 10 ppm (Table S1). These compounds were uniformly applied to the spinach samples to mimic contamination. Following this process, the samples were air-dried to enhance the binding of the analytes to the food matrix [23]. Next, each treated spinach sample was transferred into a vial, to which an equal volume of deionized water was added to maintain constant concentration levels. This step facilitated the extraction of analytes from the food matrix into the aqueous phase, enhancing their interaction with the synthesized Au@Ag NPs during SERS analysis. The concentration range of 0.5 to 10 ppm was methodically applied across five samples to generate a comprehensive dataset for analysis. Control samples were also prepared, consisting of spinach without any pesticide residues. The collection of data included the analysis of both food and water samples containing pesticides, with pesticide-free food samples serving as controls [24].

2.4. Data collection with the Raman spectroscopy

The food samples were mixed with synthesized Au@Ag NPs, and a 10 μL of the mixture was dispensed onto a gold-coated slide that was analyzed using a DXR2 Raman spectrometer (ThermoFisher Inc, Waltham, MA, USA), scanning across a wave number range from 500 to 2000 cm^{-1} with a laser power of 20 mW [22]. Representative spectra were acquired from a minimum of 300 points within each sample using OMNIC software (ThermoFisher). Noise reduction was crucial, with a 5% threshold applied to eliminate unwanted noises while preserving spectral intensity. This approach improved the signal-to-noise ratio to accurately identify pesticide residues [25]. Additionally, baseline

correction was also conducted using a quadratic polynomial equation to mitigate potential variations, ensuring an accurate depiction of spectral features. Such pre-processing improved the quality of the Raman spectral data, which is crucial for subsequent phases of analysis and interpretation [26].

2.5. Data-feature extraction and preprocessing

SERSFormer undertakes both qualitative and quantitative analyses using a unified model, utilizing preprocessed data after baseline correction. This model analyzes the intensity of individual spectra for each pesticide and concentration across consistent wavenumber ranges. Fig. 1 shows the SERS spectra of all five pesticides and the control sample, with intensity on the vertical axis (y-axis) and Raman shift on the horizontal axis (x-axis).

In qualitative analysis, the model classifies six categories, encompassing five selected pesticides and a control group. For quantitative analysis, it employs regression to predict pesticide concentrations. The SERSFormer model features two branches: classification and regression, delivering detailed outcomes on both pesticide identity and its concentration. Preprocessing, especially normalization, is crucial in machine learning to enhance model performance and interpretability. Normalization is essential for establishing a standardized scale across features, alleviating the impact of magnitude variations and mitigating issues related to exploding gradient problems during the learning process [27]. Pesticides exhibit unique Raman scattering signals, forming distinct peaks for identification. Consequently, log-min-max normalization was

applied to the Raman spectra sample intensities, ensuring a standardized scale across the dataset. The details about the log-min-max normalization are provided in the [Supplementary Material](#).

The concentration of pesticides in SERS spectra directly correlates with Raman spectra intensity levels, with higher intensities suggesting greater concentrations. As a result, regression was employed to predict concentrations. Each task requires the normalization procedures that suit the prediction task at hand. The log-min-max normalization, which was utilized in the classification task, cannot be applied to the regression task due to the potential dependency of concentration on intensity values. Min-Max normalization scales each sample to a range of [0,1], which could potentially blur the distinguishable properties of the intensities that are crucial for regression. Therefore, the log-convolution of the spectrum was employed, considering it as a time-dependent signal. Convolution, a widely used signal-processing technique, is employed to track the impact of causal input on the current input [28]. Convolution is essentially a moving windowed average. In this study, 1D Convolution was performed with a windowed Hann pulse of length 32 on the spectrum signal. The details about the convolution are provided in the [Supplementary Material](#).

Each log-convoluted spectrum sample preserves the original distinguishable intensity differences while scaling down the intensities to a trainable range for SERSFormer. In addition to the convoluted signal, we also extracted features greater than or equal to the 95, 85, 75, and 50 percentile of the intensities as additional features to be concatenated to the regression head at the later stage of the regression model. These percentile intensities were also verified based on the characteristic peaks

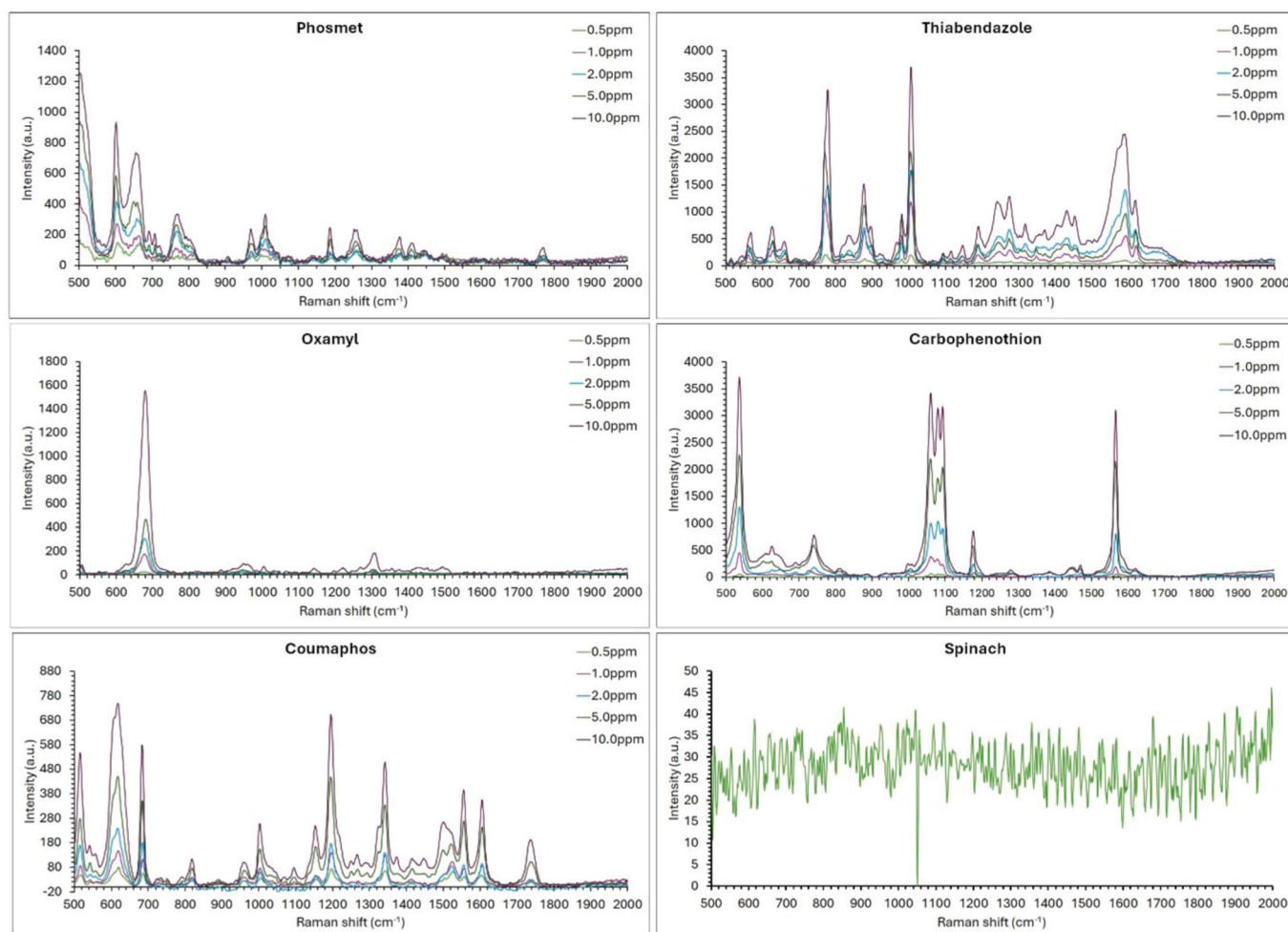


Fig. 1. SERS spectra of five pesticides and the control spinach samples.

observed in the analysis of each pesticide. Initially, the log-convoluted spectrum signal was input into the regression embedding layer of the SERSFormer.

2.6. SERSFormer model

The schematic representation of the comprehensive architecture of SERSFormer is depicted in Fig. 2. Our model comprises three main components: a task-specific embedding layer, a multi-tasking weight-sharing transformer encoder [20], and dedicated Multilayer Perceptron (MLP) heads preceding the output layers. Notably, SERSFormer encompasses two distinct branches for Classification and Regression, each equipped with an individual convolutional neural network (CNN) embedder of a similar design. The Transformer Encoder comprises multiple Multi-head Attention Encoder layers strategically shared by both the regression and classification branches. The classification branch extends into a two-layered MLP head, predicting outcomes across six distinct classes. Conversely, the Regression branch augments the contextual information regarding pesticide types by incorporating features from the Classification branch. This augmented information is then channeled into a Regression MLP head, supplemented by the concatenation of additional percentile features, yielding predictions related to the concentration of the pesticide present. The shared weights of the Transformer encoder facilitate the learning of nuanced associations, contributing to the model's dual functionality in handling both classification and regression tasks.

2.7. The embedding feature generator

The embedding layer in a deep neural network is crucial as it bridges the raw input data and the subsequent layers of the network. This transformation enables the network to identify patterns and relationships better, improving generalization and performance by providing a nuanced understanding of the data, thereby enhancing the model's accuracy and efficiency. For both regression and classification in our model, we implemented 2-layer 1D CNNs [29] with a shared stride of 3,

Relu activation, and Maxpooling on each CNN layer. CNNs can capture and leverage the contextual relationships between adjacent inputs, which is utilized in the SERSFormer architecture by retaining and using information about the influences of preceding inputs on the current ones. Consequently, the utilization of CNNs obviates the necessity for additional positional encoding in the transformer component of the SERSFormer. The 1D convolutional operations enable the network to capture hierarchical features and learn representations essential for discriminating subtle variations in the spectral data [30]. The CNN's role as an embedding layer eliminates the positional encoding overhead and more effectively understands complex patterns in the SERS data, ultimately improving performance across both task types.

2.8. Transformer encoder

The Transformer signifies a crucial advancement in neural network design, tailored for sequence transformation tasks, notably in natural language processing [11]. The Transformer deviates from the traditional recurrent or convolutional networks by utilizing a self-attention mechanism to identify intricate patterns and dependencies in sequential data. It can capture long-range dependencies without the constraints of sequential processing, thereby mitigating challenges associated with the vanishing or exploding gradients. The architecture of the Transformer, which comprises self-attention layers and feedforward sub-layers, enables parallelization, thus enhancing efficiency in both training and inference. Our model, the SERSFormer, exemplifies the Transformer Encoder, integrating a transformer attention layer. It's structured around multi-head self-attention layers [11]. The attention mechanism empowers the model to selectively emphasize different segments of the input sequence during the prediction process. Within this mechanism, the concept of multi-head attention is employed, wherein multiple sets, or "heads," independently perform scaled dot-product self-attention calculations. The scaled dot-product self-attention computation involves utilizing query, key, and value matrices to derive attention scores, subsequently guiding the weighted summation of values.

Fig. 3 explains the multi-head scaled dot product self-attention layer

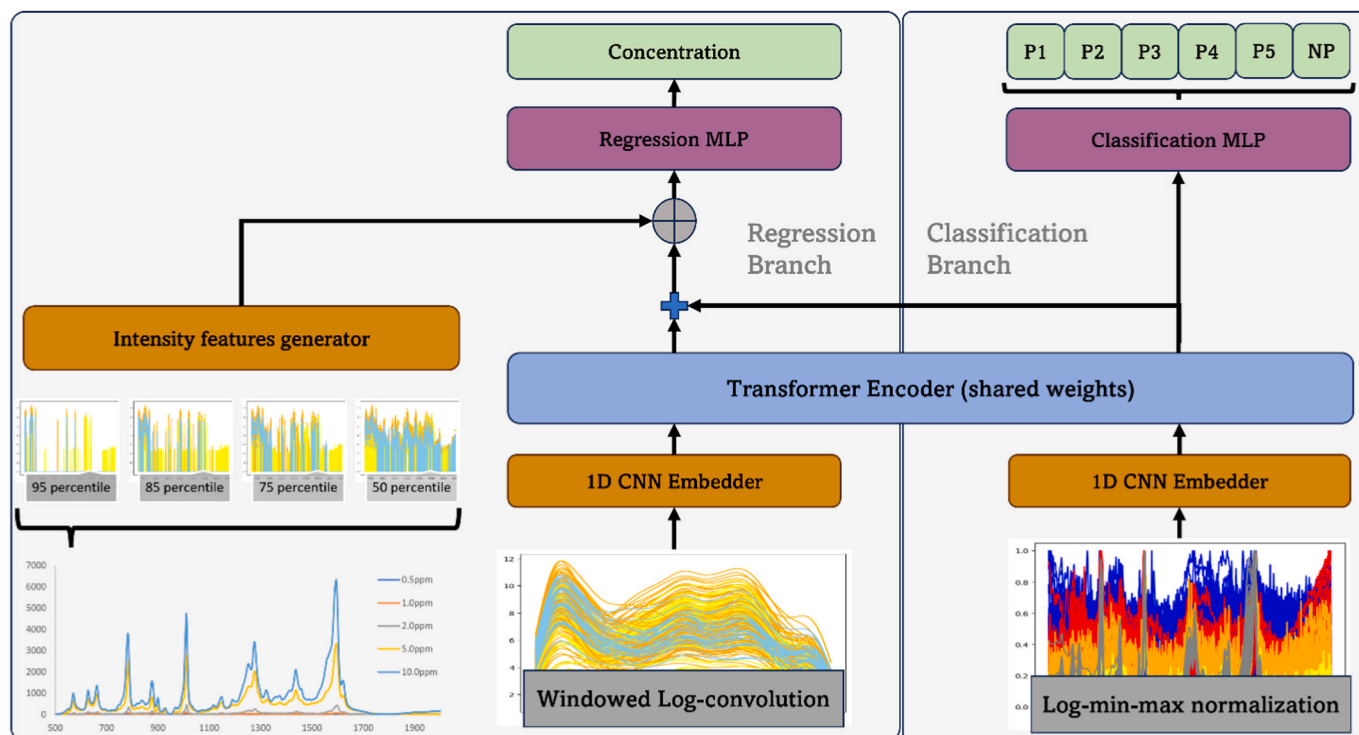


Fig. 2. Overview of SERSFormer architecture that uses shared weights Transformer Encoder for classification (qualitative) and regression (quantitative) tasks.

of the transformer encoder module. The important steps in the multi-head attention layer are provided in the [Supplementary Materials](#). This configuration enhances the model's capacity to discern and weigh the relevance of different elements within the input sequence, thereby optimizing the encoding process. The nuanced derivation of attention within the SERSFormer model underscores its theoretical foundation rooted in the transformative concepts introduced in the pioneering work on attention mechanisms in neural networks [11].

Both classification and regression tasks employ a six-layer transformer encoder with four attention heads with shared weights, allowing the Transformer encoder to discern intricate relationships between various types of pesticides and their respective concentrations. For classification, the output from the transformer encoder is directly utilized as input for the Classification Multi-layer Perceptron (MLP) module with two hidden linear layers, Rectified Linear Unit (ReLU) activation functions, and batch normalization, ultimately yielding six outputs with softmax activation for the generation of multi-class classification probabilities. In regression, before being fed into the regression MLP head, a skip connection is established by adding the encoded features from the transformer encoder of the classification branch to those of the regression branch. The summation of values is performed to maintain dimensionality consistency. Subsequently, the augmented features, including additional percentile features containing information about the magnitude of the input signals, are concatenated with the newly encoded features. This expanded feature set is then fed into the regression MLP module, which consists of two hidden layers with ReLU activation functions and batch normalization, aiming to predict concentrations accurately.

2.9. Training of data

The dataset is partitioned into training, validation, and test sets with a 70:10:20 ratio to facilitate proper model evaluation. While splitting, it's ensured that all the classes and all the concentration values in the test dataset are in equal ratio and are unseen to the training to ensure rigorous testing. For classification tasks, the model employs the cross-entropy loss function, Adam optimizer, with a learning rate of 0.0001. The architecture encompasses 4 attention heads and 6 layers of a Transformer encoder, enriched with regularization techniques such as dropouts of 0.5 and batch normalization in between every embedding layer and MLP layer. The output layer utilizes softmax activation to facilitate the classification decision. Conversely, for regression tasks, the mean squared error loss function is adopted, coupled with the Adam

optimizer employing a learning rate of 0.0001. The regression-oriented Transformer model mirrors the architecture utilized in classification, featuring 4 attention heads and 6 layers, while regularization is achieved through the incorporation of dropouts of 0.5 and batch normalization between every CNN layer and MLP layer. The complete model undergoes 100 epochs of training, with early stopping privilege. The table below shows the best hyperparameters for the best performance of the SERSFormer ([Table 1](#)).

3. Results & discussion

3.1. Characterization of synthesized nanoparticles via TEM-EDS

The synthesized Au@Ag NPs were characterized using TEM-EDS, revealing a well-defined nanoparticle structure with an average diameter of 21.44 ± 4.02 nm. These nanoparticles exhibited a normal dispersion, indicating a uniform distribution across the sample. TEM images unveiled spherical nanoparticles, while EDS elemental mapping confirmed the core-shell configuration ([Fig. 4](#)). Specifically, gold was identified, forming the core, encased by a distinct silver layer serving as the shell. This characterization confirmed the successful synthesis of the core-shell nanoparticles and provided insights into their size, morphology, and elemental composition, affirming their suitability for subsequent SERS analysis.

3.2. Characteristic Raman peaks of pesticide molecules

This study analyzed five distinct pesticide samples—coumaphos, oxamyl, carbophenothion, thiabendazole, and phosmet—each displaying characteristic patterns in their Raman spectral data. These patterns can be used for identifying pesticide residues by chemometric techniques. For instance, coumaphos, an organophosphorus pesticide, exhibits distinctive peaks at 519, 616, 648, 1197, 1343, 1556, 1607, and 1740 cm^{-1} in its Raman spectrum. These peaks correspond to specific vibrational modes related to its molecular structure, such as P-C stretching, P-O-C bending, C-H bending, C-N stretching, and C=O stretching vibrations [31]. Similarly, oxamyl, a carbamate pesticide, displays prominent peaks at 681 and 1311 cm^{-1} , indicative of C-N bending and C-N stretching vibrations, respectively [32].

Similarly, carbophenothion, another organophosphorus pesticide, displays distinct peaks at 536, 1059, 1078, 1093, 1174, and 1566 cm^{-1} in its Raman spectrum, corresponding to various vibrational modes such as P = S stretching, P-O-C stretching, and P-S-C bending vibrations [31].

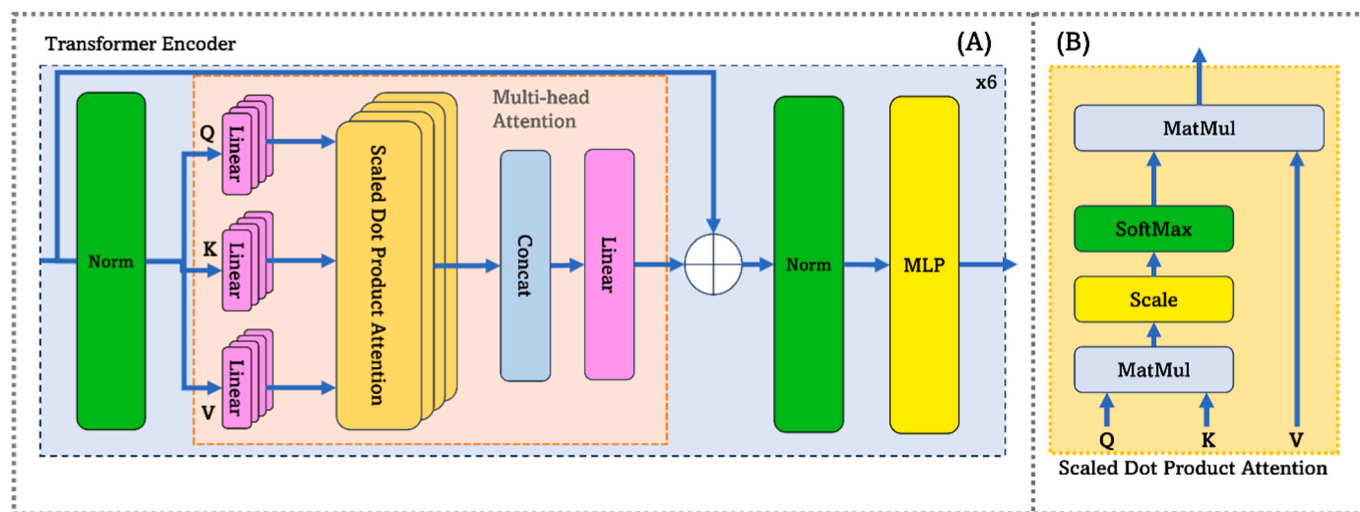
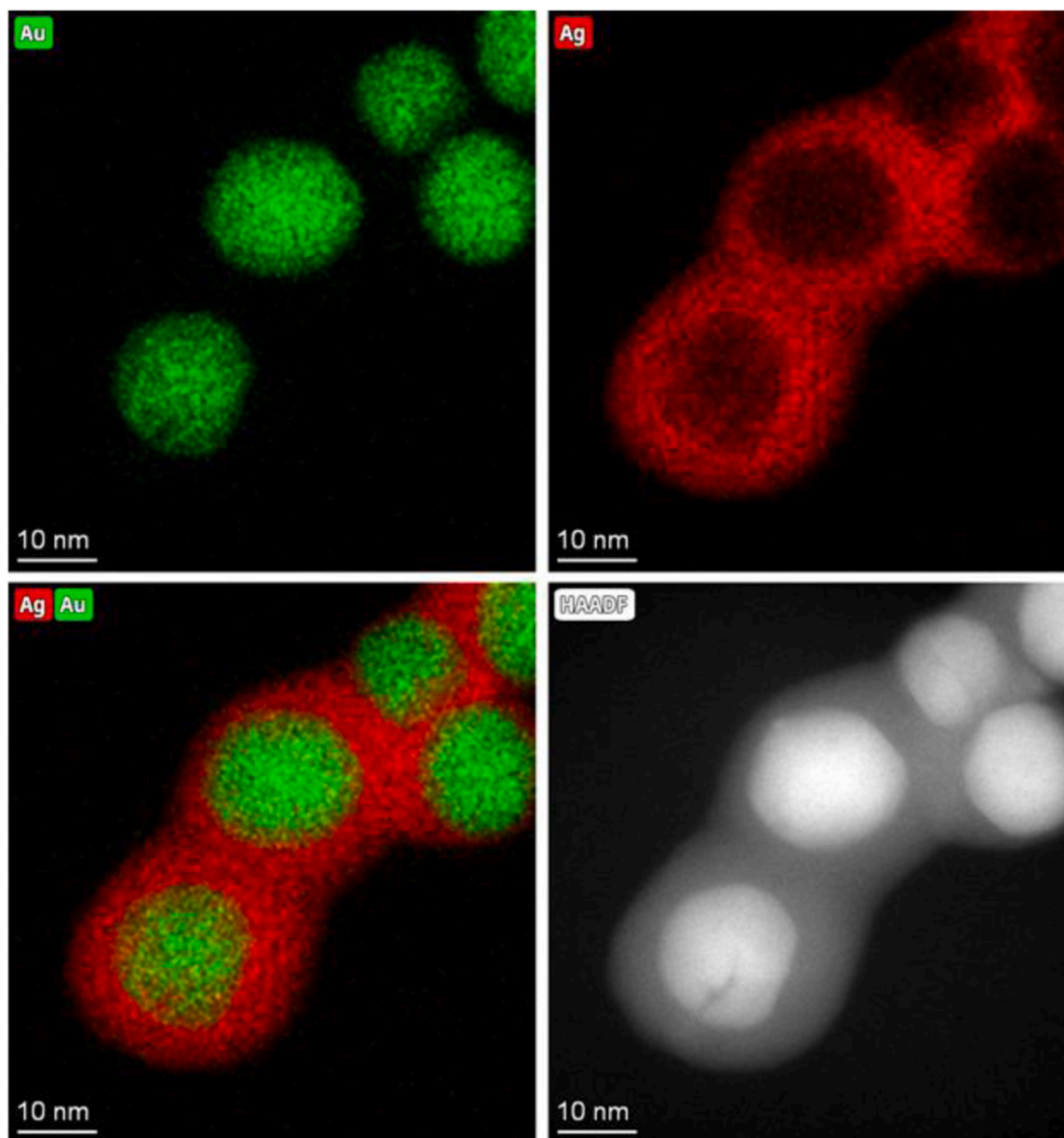


Fig. 3. Design of multi-head (4-head) transformer encoder module with six scaled dot product attention layers (A); block diagram of the scaled dot product attention mechanism (B).

Table 1

The best hyperparameters for the best performance of the SERSFormer.

Name	Learning rate	Attention head	Encoder layers	Number of classes	Epochs	Batch size
SERSFormer	0.0001	4	6	6	45	32

**Fig. 4.** TEM-EDS analysis of synthesized gold-silver core-shell nanoparticles.

Thiabendazole, a benzimidazole fungicide, manifests characteristic peaks at 771, 883, 981, 1001, 1592, and 1626 cm^{-1} , representing C-S stretching, C-N stretching, C-H in-plane bending, and C=C stretching vibrations [22]. Additionally, phosmet, an organophosphorus pesticide, exhibits distinct features in its Raman spectrum, with peaks at 602, 646, 765, 1013, and 1174 cm^{-1} , corresponding to P-S stretching, P-O-C stretching, P-O-C bending, and P-S-C bending vibrations [33]. The unique spectral profiles of these pesticides facilitate their identification and enable the development of machine learning models for accurately discerning and quantifying their presence in samples.

3.3. Evaluation and results of SERSFormer

SERSFormer integrates the functionalities of both classification and regression to deliver thorough qualitative and quantitative assessments. Each SERS spectrum sample analyzed by SERSFormer yields two concurrent outcomes: identification of the pesticide type, if present, and determination of its concentration. In cases where no pesticide is found, the model assigns a concentration value of 0 ppm. To evaluate the model's efficacy, we conducted separate evaluations for the classification and regression tasks. For the multiclass classification, the evaluation of model's performance requires the use of various metrics. This study employed multiclass precision, recall, accuracy, and F1 score,

each providing unique insights into the classifier's efficiency [34].

- Precision, defined as the ratio of true positive predictions to the sum of true positives and false positives, elucidates the model's accuracy in correctly identifying instances of a specific class.
- Recall, in contrast, measures the model's capability to capture all instances of a particular class by measuring the ratio of true positives to the sum of true positives and false negatives.
- Accuracy, a fundamental metric, gauges the overall correctness of the classification model by calculating the ratio of correctly predicted instances to the total number of instances.
- The F1 score, a harmonic mean of precision and recall, combines these metrics into a single value, providing a balanced assessment of the model's performance.

Our SERSFormer achieves a 98% accuracy rate in identifying and classifying pesticides, as evidenced in Table 2. All evaluation metrics confirm that SERSFormer proficiently conducts qualitative analyses of food contamination.

All these evaluation metrics can be derived from the confusion matrix, a pivotal tool in assessing the performance of classification models, particularly in multiclass scenarios (Fig. 5). The confusion matrix is a table that displays classification outcomes by comparing predicted labels against true labels across various classes. In a multiclass setting, this matrix takes the form of a square, with each row representing the true class and each column representing the predicted class. The diagonal elements of the matrix represent the instances that are correctly classified, while off-diagonal elements indicate misclassifications. The six classes represent five distinct pesticides (thiabendazole, phosmet, carbophenothion, coumaphos, and oxamyl, respectively) and one class for samples without any pesticides. Accordingly, the confusion matrix is a 6×6 matrix. Each cell (i, j) within this matrix denotes the number of instances from class i that were predicted as class j . For instance, if cell $(3, 3)$ holds the value 258, it means that 258 instances belonging to class 3 (which corresponds to a specific pesticide) were correctly predicted as class 3. Conversely, if cell $(4, 5)$ contains the value 1, it means that one instance from class 4 was misclassified as class 5. By examining the confusion matrix, various performance metrics such as precision, recall, accuracy, and the F1 score can be calculated for each class, providing a detailed and nuanced assessment of the model's ability to distinguish among the different classes in a multiclass classification problem. The confusion matrix is an indispensable tool for understanding the strengths and weaknesses of a model across diverse classes, thereby aiding in the refinement and optimization of classification algorithms. The confusion matrix from the test dataset corroborates the high efficiency of our evaluation metrics.

For regression of concentration, evaluating predictive models is crucial, and several metrics serve as benchmarks to assess the model's performance. We use the R^2 score, mean absolute error (MAE), and mean squared error (MSE) as evaluation metrics [35]. The MAE quantifies the average magnitude of errors between the predicted and actual values, providing a direct metric of accuracy. While the MSE calculates the average of squared differences between predicted and actual values, giving more weight to larger errors. Both MAE and MSE offer insights into the precision of the model's predictions, with lower values signifying superior accuracy. This triad of metrics collectively affords a holistic evaluation of the regression model's ability to accurately predict concentrations, considering both the overall fit to the data and the

magnitude of predictive errors. Such comprehensive assessments are integral to refining and optimizing regression models for enhanced predictive capabilities in diverse applications. The table below shows the evaluation metrics for the quantitative analysis of pesticides on the test dataset (Table 2).

Fig. 6 depicts the predictive analysis of pesticide concentration, with the predicted pesticide concentration plotted against the actual concentrations. Due to the variability in the input sample spectrum within the concentration, there are possibilities of outliers in the regression. With improved sample data and a pool of concentrations, we plan to improve the SERSFormer further (Fig. 6).

3.4. Kernel density estimates of regression

Fig. 7 presents kernel density violin plots, a graphical tool that combines elements of box plots and kernel density estimation (KDE) to provide a detailed representation of the density distribution in quantitative predictions. These plots effectively convey density characteristics, with the highest bulge indicating the region of maximum density. The precise predicted density is visualized using grey sticks. Violin plots serve as a visual aid in illustrating the distribution of predicted concentrations by utilizing the width of the plot at specific concentration levels, thereby encapsulating the estimated density [36]. In Fig. 7, the X-axis represents the actual concentration, and the Y-axis represents the predicted concentration from SERSFormer on the test dataset. The various targets along the X-axis denote concentrations of 0, 0.5, 1, 2, 5, and 10 ppm, respectively. The KDE estimate clearly demonstrates the ability of SERSFormer to predict concentrations, as the highest density closely aligns with the actual concentration levels. Additionally, the model demonstrates a noteworthy proficiency in predicting uncontaminated samples without any pesticide residues. The decrease in prediction efficacy at concentrations of 0.5 and 1 ppm can be attributed to subtle differences in intensity levels. The intensity levels are directly associated with pesticide concentrations, as detailed in the Discussion section (refer to Fig. 6 and Table 3 for more details). Thus, the kernel density violin plots serve as an analytical tool that facilitates a nuanced understanding of the precision and distribution of SERSFormer predictions across a range of concentration levels.

3.5. Importance of shared weights

The utilization of multi-tasking weight sharing in the SERSFormer model, namely employing a Transformer encoder for both classification and regression tasks, offer several significant advantages. The adoption of weight-sharing enhances parameter efficiency, enabling the model to acquire a succinct representation of input data suitable for both tasks. This is particularly beneficial in situations where data availability, such as SERS data for both tasks, is limited. Furthermore, shared weights act as a mechanism for transfer learning, allowing the seamless transfer of knowledge from classification tasks to regression tasks and vice versa, thereby enhancing performance across the board [20]. Additionally, the multi-task weight sharing drastically decreases the overall complexity of the model. Instead of maintaining separate models for classification and regression, the shared parameterization simplifies the architecture, leading to expedited training and inference times. In regression, shared weights promote the reusability of features, empowering the model to extract pertinent features that capture underlying patterns or relationships in the data. This shared knowledge enhances the model's accuracy

Table 2

Evaluation metrics for the test dataset for classification and quantitative analysis of the pesticides on the test dataset.

Evaluation metrics for classification				Evaluation metrics for quantitative/regression		
Multiclass accuracy	Multiclass F1 score	Multiclass precision	Multiclass recall	Mean absolute error	Mean squared error	R^2 score
0.984	0.982	0.983	0.984	0.966	1.826	0.849

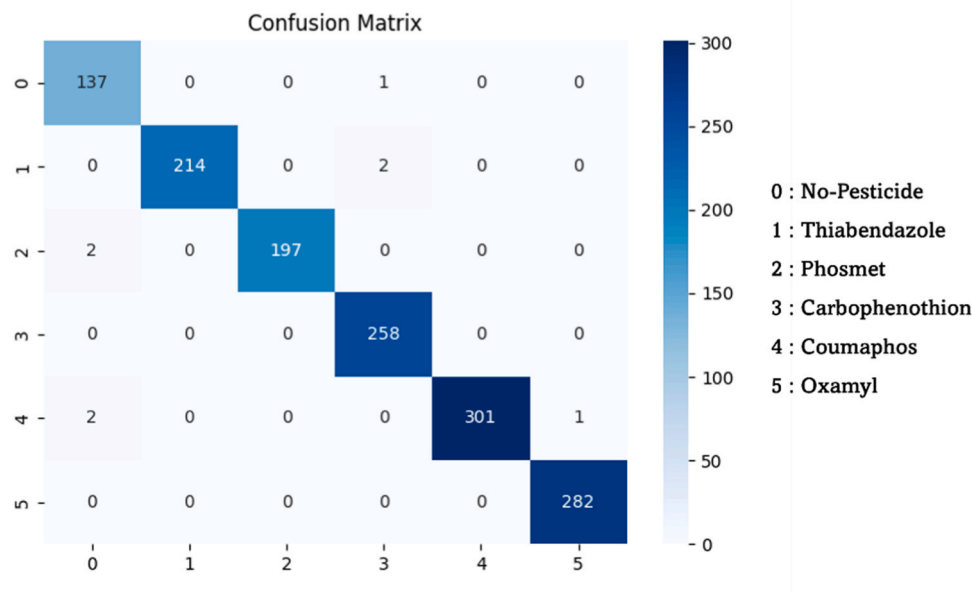


Fig. 5. Confusion Matrix for identification of pure spinach sample (class 0) and five different pesticides (thiabendazole, phosmet, carbophenothion, coumaphos, and oxamyl, respectively).

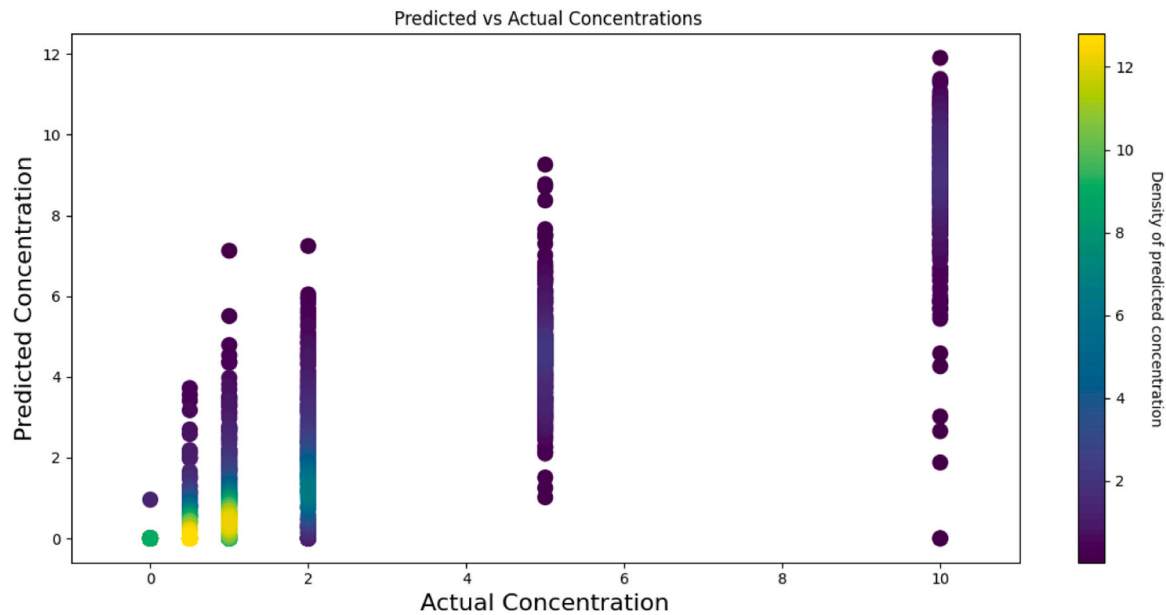


Fig. 6. The predictive analysis of the concentration of pesticides, with predicted pesticide concentrations plotted against true concentrations.

in predicting continuous values. Moreover, the SERSFormer model, through the utilization of shared weights, demonstrates improved generalization capabilities. The model acquires a more comprehensive representation of input features that surpasses the specific requirements of classification or regression tasks, thus fostering enhanced its adaptability to new and unseen data. Shared weights promote a unified learning approach, enabling the model to dynamically adjust its parameters based on the intrinsic characteristics of the data, whether for classification or regression. This amalgamation of benefits underscores the efficacy of multi-task weight sharing in the SERSFormer model, providing a versatile and efficient solution for tackling both classification and regression objectives.

3.6. Effects of different normalization procedures on classification

In this ablation study, our objective was to scrutinize the pivotal factors influencing qualitative prediction and emphasize the significance of efficient signal or data processing in deep learning models. The SERSFormer model underwent training with distinct data processing techniques aimed at comprehending the model's learning behavior and identifying the key factors influencing its efficiency in classification tasks. Various normalization methods, including min-max normalization, log-min-max normalization, z-score normalization, and convoluted signals of various window lengths used in regression features, were employed [28,37]. Fig. 8 illustrates a sample spectrum of spinach sample contaminated with oxamyl, processed using different normalization techniques.

Table 3 presents the outcomes of these data processing techniques on

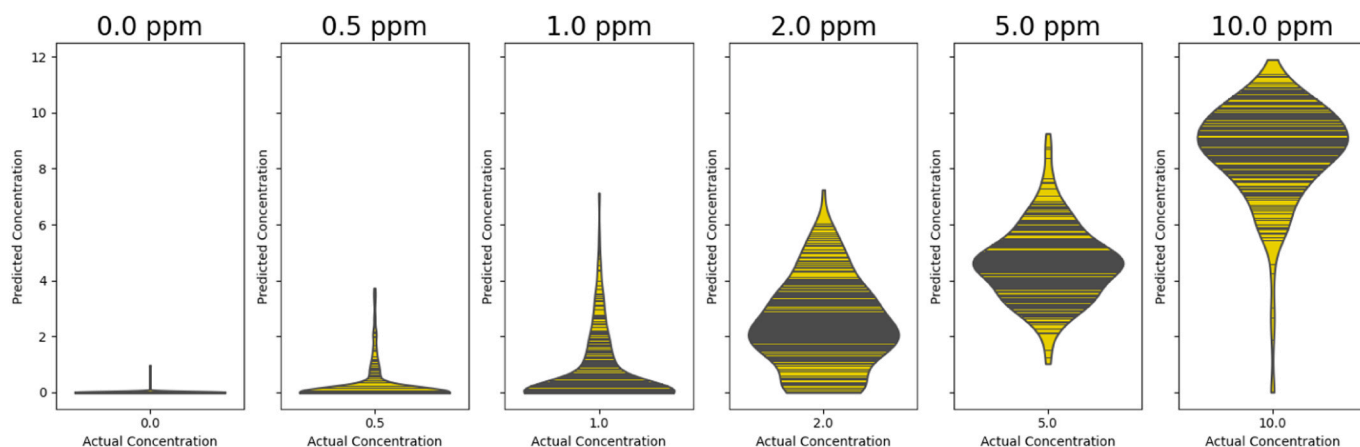


Fig. 7. Kernel density violins for each predicted concentration against true concentration.

Table 3

Effects of different data-processing techniques on the classification task of SERSFormer.

Normalization types/Metrics	Accuracy	Precision	Recall	F1-Score
Min-max normalization	0.967	0.964	0.967	0.963
Log-min-max normalization	0.984	0.983	0.984	0.982
Z-score normalization	0.965	0.957	0.965	0.958
Log-convolution-389	0.615	0.6638	0.615	0.554
Log-convolution-32	0.956	0.946	0.956	0.941

model's performance. Notably, log-min-max normalization displayed the best performance, followed by min-max normalization and log-z-score normalization. Conversely, the log-convoluted signal with a window length of 389 exhibited the least favorable performance. Scientifically, it is widely accepted that each pesticide has unique functional groups and molecular properties, which result in distinct Raman shift

peaks at specific wavenumbers within the spectrum. This experiment provides further evidence for these scientific principles, confirming that our model effectively captures the intricate relationship between Raman peaks and the types of pesticides. The suboptimal performance of the convolution-389 signal can be attributed to its tendency to smooth out exact peak positions, leading to the loss of minute details in the signals. While min-max normalization performs better than convoluted signals, it may sometimes cause the model to neglect smaller peaks, especially when the differences between peaks are substantial, resulting in diminished gradients and suboptimal training. Log-min-max normalization mitigates this issue by amplifying smaller peaks and providing adequate information within the necessary normalization range, thereby contributing to more effective model training.

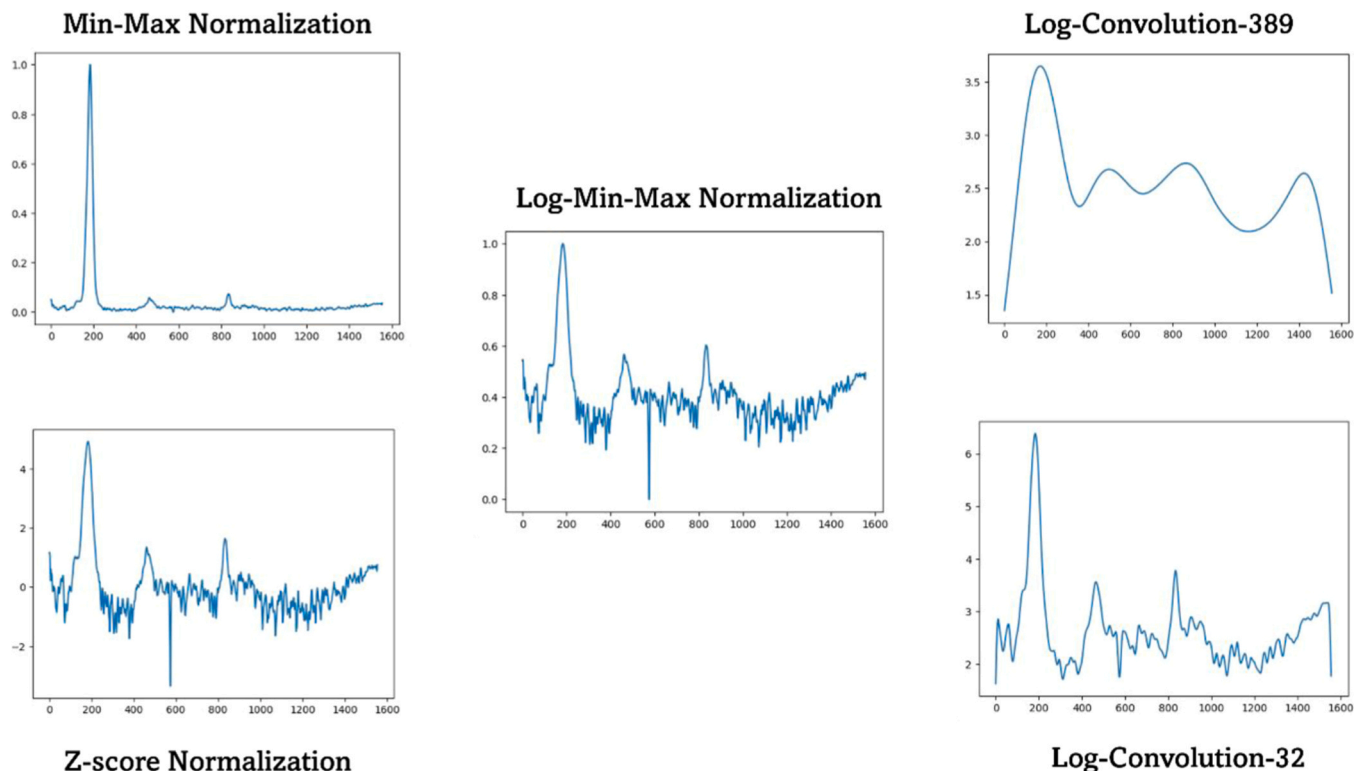


Fig. 8. Plots of data processing techniques applied on the spectra of oxamyl samples for a classification task.

3.7. Influence of convolution on regression

This section provides an analysis of the impact of various normalization and data processing techniques on the prediction of analyte concentrations. Our investigation involved the application of min-max normalization, utilized in the classification branch, and convolutions with different window lengths on regression features. The objective was to ascertain whether the model could effectively learn inherent regression features associated with concentration and to identify critical features influencing concentrations. In Fig. 9, it is observed that both min-max and log-min-max normalizations scale down concentrations to the [0,1] range, resulting in uniform intensities across all concentrations. This uniformity potentially hampers the model's learning process, making it challenging to distinguish between different concentration inputs. In contrast, the convoluted signal demonstrates the ability to differentiate between two distinct input concentrations. Convolution, a widely employed processing technique in Digital Signal Processing, functions as a causal moving average technique. It effectively reduces higher variability in the data while preserving sufficient differences between signal amplitudes [38].

Table 4 elucidates the effects of various input processing techniques on the model's performance in regression tasks. As anticipated, the model normalized with min-max struggles to learn, resulting in poor performance, whereas models with convoluted inputs exhibit significantly higher efficiency. The Regression branch leverages shared weights from the Transformer and integrates context regarding the type of pesticide from the Classification branch, which inherently relies on the positions of wavenumbers. In this context, Convolution 32 emerges as the most effective, as it retains crucial information about minute peak positions needed for classifying pesticide types while effectively distinguishing between different concentrations. These findings substantiate the assertion that the quantity of pesticide in a sample is determined by the intensity levels in the SERS spectrum, specific to each sample. Consequently, our SERSFormer demonstrates robust regression

Table 4

Effects of different data-processing techniques on regression task of SERSFormer.

Regression features/Metrics	MAE	MSE	R ² -Score
Min-max normalization	1.379	5.534	-13.127
Log-convolution-389	1.165	3.371	0.6598
Log-convolution-64	1.632	4.484	0.5648
Log-convolution-32	0.966	1.826	0.849

capabilities, accurately predicting concentrations based on the distinct characteristics of each pesticide.

While the proposed methodology shows promising results, several limitations require consideration. The reliance on synthesized Au@Ag NPs introduces a potential source of variability. Although efforts were made to ensure daily synthesis and characterization, subtle variations may still occur over time, affecting the reproducibility of results. Future research should explore strategies for nanoparticle stabilization to minimize these temporal variations. Additionally, the performance of the current models might be influenced by the specificity of the pesticides analyzed. Expansion of the analyte set to include a broader range of pesticides would enhance the model's applicability in diverse agricultural settings. Furthermore, incorporating data from real-world scenarios with varying environmental conditions and sample matrices will further validate the model's robustness.

4. Conclusions

This study synergizes cutting-edge machine learning models and advanced SERS techniques to rapidly and accurately detect pesticide residues in agricultural products. Integrating SERS and the SERSFormer model demonstrates the potential to transform pesticide analysis with high sensitivity, specificity, and efficiency. Gold-silver core-shell nanoparticles, serving as SERS substrates, ensure significant enhancement of Raman scattering signals. TEM-EDS analysis confirms successful

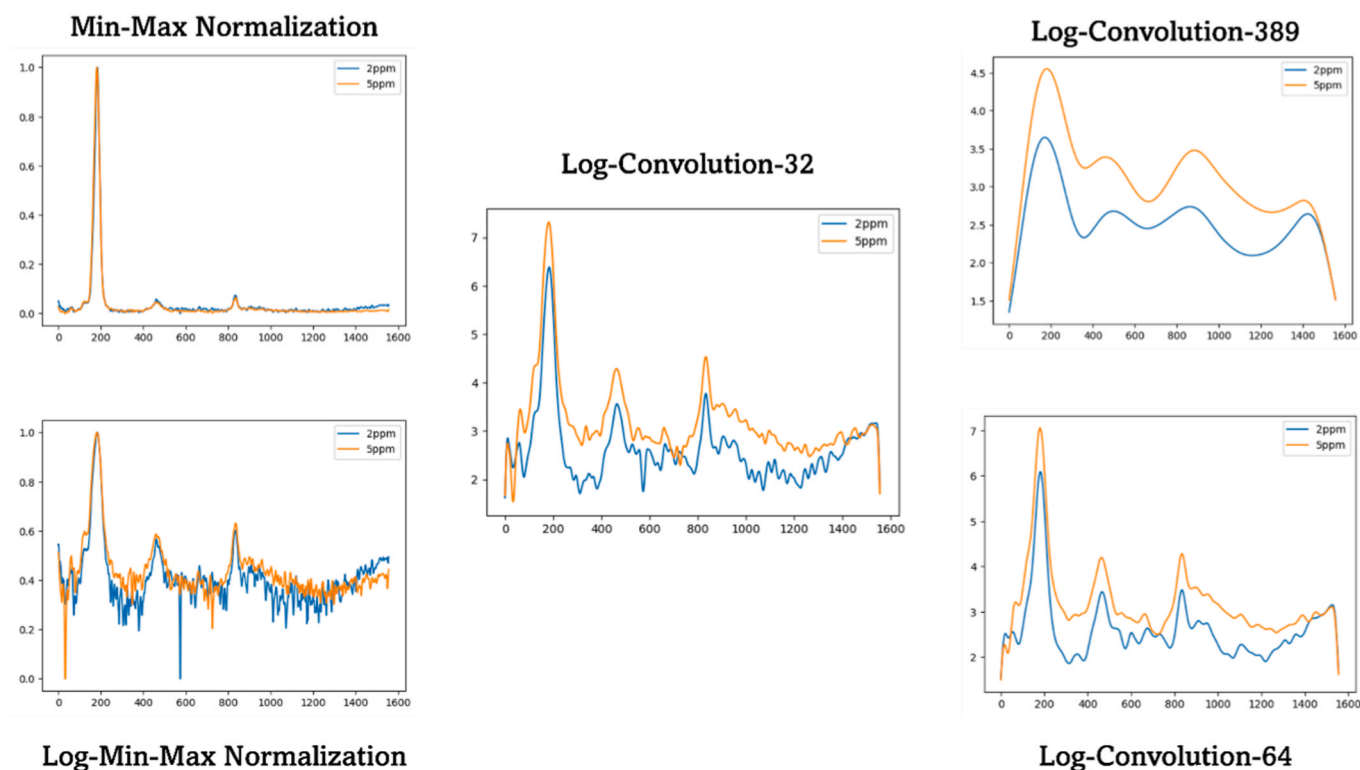


Fig. 9. Plots illustrating data processing techniques applied on the spectra of oxamyl samples for a regression task, using two different concentrations: 2 and 5 ppm, respectively.

nanoparticle synthesis and provides insights into their structure and composition. The SERSFormer serves as a versatile tool for both qualitative and quantitative analysis. Qualitatively, it accurately identified six pesticide categories, benefiting significantly from preprocessing techniques like noise reduction, baseline correction, and normalization. Quantitatively, the model excelled in predicting pesticide concentrations, with the integration of log-convolution and percentile feature extraction capturing subtle concentration-dependent spectral features. Evaluation metrics such as R^2 Score, MAE, and MSE highlighted the model's quantitative accuracy, offering a comprehensive assessment of its predictive capabilities. The study underscores the significance of data normalization techniques in pesticide classification and quantification tasks, including log-min-max normalization, log-convolution, and confusion matrix analysis. This integrated approach, combining SERS with machine learning, offers a promising route for rapid, reliable pesticide detection, with significant implications for monitoring food safety in the agriculture and food sectors.

Environmental implication statement

Pesticides, known for their detrimental impact on ecosystems, human health, and non-target organisms, persist in the environment, contaminating soil, water, and food. Our development of SERS-based rapid pesticide detection coupled with machine learning tackles this crucial issue. By swiftly and accurately identifying pesticide residues in food, our method promotes food safety and encourages sustainable agricultural practices. This approach safeguards ecosystems and public health by enabling early detection and monitoring of pesticide contamination.

CRediT authorship contribution statement

Mengshi Lin: Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Akshata Hegde:** Writing – original draft, Methodology, Formal analysis. **Mehdi Hajikhani:** Writing – original draft, Methodology, Formal analysis, Data curation. **Jianlin Cheng:** Writing – review & editing, Funding acquisition, Conceptualization. **John Snyder:** Writing – review & editing, Validation, Formal analysis.

Declaration of Competing Interest

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

Data availability

Data will be made available on request.

Acknowledgements

This research was financially supported by the USDA National Institute of Food and Agriculture (2020–67030-31336 & 2023–67017-40165) and the National Science Foundation (CBET-2103025).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.jhazmat.2024.134208](https://doi.org/10.1016/j.jhazmat.2024.134208).

References

- [1] Leong, W.H., Teh, S.Y., Hossain, M.M., Nadarajaw, T., Zabidi-Hussin, Z., Chin, S.Y., Lai, K.S., Lim, S.E., 2020. Application, monitoring and adverse effects in pesticide use: The importance of reinforcement of Good Agricultural Practices (GAPs). *J Environ Manag* 260, 109987. <https://doi.org/10.1016/j.jenvman.2019.109987>.
- [2] Sun, X., Zhao, Y., Liu, L., Qiao, Y., Yang, C., Wang, X., Li, Q., Li, Y., 2024. Visual whole-process monitoring of pesticide residues: An environmental perspective using surface-enhanced Raman spectroscopy with dynamic borohydride-reduced silver nanoparticles. *J Hazard Mater* 465, 133338. <https://doi.org/10.1016/j.jhazmat.2023.133338>.
- [3] Damalas, C.A., Koutroubas, S.D., 2016. Farmers' Exposure to Pesticides: Toxicity Types and Ways of Prevention. *Toxics* 4 (1), 1. <https://doi.org/10.3390/toxics4010001>.
- [4] Bhatt, P., Zhou, X., Huang, Y., Zhang, W., Chen, S., 2021. Characterization of the role of esterases in the biodegradation of organophosphate, carbamate, and pyrethroid pesticides. *J Hazard Mater* 411, 125026. <https://doi.org/10.1016/j.jhazmat.2020.125026>.
- [5] Zhou, X., Liu, G., Zhang, H., Li, Y., Cai, W., 2019. Porous zeolite imidazole framework-wrapped urchin-like Au-Ag nanocrystals for SERS detection of trace hexachlorocyclohexane pesticides via efficient enrichment. *J Hazard Mater* 368, 429–435. <https://doi.org/10.1016/j.jhazmat.2019.01.070>.
- [6] Wang, T., Wang, S., Cheng, Z., Wei, J., Yang, L., Zhong, Z., Hu, H., Wang, Y., Zhou, B., Li, P., 2021. Emerging core-shell nanostructures for surface-enhanced Raman scattering (SERS) detection of pesticide residues. *Chem Eng J* 424, 130323. <https://doi.org/10.1016/j.cej.2021.130323>.
- [7] Pérez-Jiménez, A.I., Lyu, D., Lu, Z., Liu, G., Ren, B., 2020. Surface-enhanced Raman spectroscopy: benefits, trade-offs and future developments. *Chem Sci* 11 (18), 4563–4577. <https://doi.org/10.1039/D0SC00809E>.
- [8] Xu, M.-L., Gao, Y., Han, X.-X., Zhao, B., 2022. Innovative Application of SERS in Food Quality and Safety: A Brief Review of Recent Trends. *Foods* 11 (14). <https://doi.org/10.3390/foods11142097>.
- [9] Moldovan, R., Iacob, B.C., Farcău, C., Bodoki, E., Oprean, R., 2021. Strategies for SERS Detection of Organochlorine Pesticides. *Nanomater (Basel)* 11 (2), 304. <https://doi.org/10.3390/nano11020304>.
- [10] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., 2020. Language models are few-shot learners. *Adv Neural Inf Process Syst* 33, 1877–1901. <https://doi.org/10.48550/arXiv.2005.14165>.
- [11] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention is all you need. *Adv Neural Inf Process Syst* 30.
- [12] D. Wang, M. Pourmirzaei, U.L. Abbas, S. Zeng, N. Manshour, F. Esmaili, B. Poudel, Y. Jiang, Q. Shao, J. Chen, D. Xu, 2024. S-PLM: Structure-aware Protein Language Model via Contrastive Learning between Sequence and Structure. *bioRxiv*. <https://doi.org/10.1101/2023.08.06.552203>.
- [13] El Naqa, I., Murphy, M.J., 2015. What is machine learning? In: El Naqa, I., Li, R., Murphy, M.J. (Eds.), *Machine Learning in Radiation Oncology: Theory and Applications*. Springer International Publishing, Cham, pp. 3–11. https://doi.org/10.1007/978-3-319-18305-3_1.
- [14] Sarker, I.H., 2021. Machine learning: algorithms, real-world applications and research directions. *SN Comput Sci* 2 (3), 160. <https://doi.org/10.1007/s42979-021-00592-x>.
- [15] Yu, F.R., He, Y., 2019. Introduction to Machine Learning. In: Yu, F.R., He, Y. (Eds.), *Deep Reinforcement Learning for Wireless Networks*. Springer International Publishing, Cham, pp. 1–13. https://doi.org/10.1007/978-3-030-10546-4_1.
- [16] Oliveira, L., Gomes Domingos, A.L., Louzada Fernandes, D., Ribeiro Cerqueira, F., Siqueira-Batista, R., Bressan, J., 2023. Applicability of machine learning techniques in food intake assessment: A systematic review. *Crit Rev Food Sci Nutr* 63 (7), 902–919. <https://doi.org/10.1080/10408398.2021.1956425>.
- [17] Khan, M.I.H., Sablani, S.S., Nayak, R., Gu, Y., 2022. Machine learning-based modeling in food processing applications: State of the art. *Compr Rev Food Sci Food Saf* 21 (2), 1409–1438. <https://doi.org/10.1111/1541-4337.12912>.
- [18] Sarker, I.H., 2021. Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. *SN Comput Sci* 2 (6), 420. <https://doi.org/10.1007/s42979-021-00815-1>.
- [19] Wang, J., Chen, Q., Belwal, T., Lin, X., Luo, Z., 2021. Insights into chemometric algorithms for quality attributes and hazards detection in foodstuffs using Raman/surface enhanced Raman spectroscopy. *Compr Rev Food Sci Food Saf* 20 (3), 2476–2507. <https://doi.org/10.1111/1541-4337.12741>.
- [20] S. Ruder, 2017. An overview of multi-task learning in deep neural networks. *arXiv* 1706, 05098. <https://doi.org/https://doi.org/10.48550/arXiv.1706.05098>.
- [21] Chen, X., Lin, M., Sun, L., Xu, T., Lai, K., Huang, M., Lin, H., 2019. Detection and quantification of carbendazim in Oolong tea by surface-enhanced Raman spectroscopy and gold nanoparticle substrates. *Food Chem* 293, 271–277. <https://doi.org/10.1016/j.foodchem.2019.04.085>.
- [22] Hajikhani, M., Kousheh, S., Zhang, Y., Lin, M., 2024. Design of a novel SERS substrate by electrospinning for the detection of thiabendazole in soy-based foods. *Food Chem* 436, 137703. <https://doi.org/10.1016/j.foodchem.2023.137703>.
- [23] Yang, T., Zhang, Z., Zhao, B., Hou, R., Kinchla, A., Clark, J.M., He, L., 2016. Real-time and in situ monitoring of pesticide penetration in edible leaves by surface-enhanced Raman scattering mapping. *Anal Chem* 88 (10), 5243–5250. <https://doi.org/10.1021/acs.analchem.6b00320>.
- [24] Kang, Y., Li, L., Chen, W., Zhang, F., Du, Y., Wu, T., 2018. Rapid In Situ SERS Analysis of Pesticide Residues on Plant Surfaces Based on Micelle Extraction of Targets and Stabilization of Ag Nanoparticle Aggregates. *Food Anal Methods* 11 (11), 3161–3169. <https://doi.org/10.1007/s12161-018-1290-2>.
- [25] Deng, J., Zhang, X., Li, M., Jiang, H., Chen, Q., 2022. Feasibility study on Raman spectra-based deep learning models for monitoring the contamination degree and level of aflatoxin B1 in edible oil. *Microchem J* 180, 107613. <https://doi.org/10.1016/j.microc.2022.107613>.

- [26] Martyna, A., Menzyk, A., Damin, A., Michalska, A., Martra, G., Alladio, E., Zadora, G., 2020. Improving discrimination of Raman spectra by optimising preprocessing strategies on the basis of the ability to refine the relationship between variance components. *Chemom Intell Lab Syst* 202, 104029. <https://doi.org/10.1016/j.chemolab.2020.104029>.
- [27] Singh, D., Singh, B., 2020. Investigating the impact of data normalization on classification performance. *Appl Soft Comput* 97, 105524. <https://doi.org/10.1016/j.asoc.2019.105524>.
- [28] Keys, R., 1981. Cubic convolution interpolation for digital image processing. *IEEE Trans Acoust, Speech, Signal Process* 29 (6), 1153–1160. <https://doi.org/10.1109/TASSP.1981.1163711>.
- [29] Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., Inman, D.J., 2021. 1D convolutional neural networks and applications: A survey. *Mech Syst Signal Process* 151, 107398. <https://doi.org/10.1016/j.ymssp.2020.107398>.
- [30] Kiranyaz, S., Ince, T., Abdeljaber, O., Avci, O., Gabbouj, M., 2019. 1-D Convolutional Neural Networks for Signal Processing Applications. *ICASSP 2019 - 2019 IEEE Int Conf Acoust, Speech Signal Process (ICASSP)* 8360–8364.
- [31] Dowgiallo, A.M., Guenther, D.A., 2019. Determination of the Limit of Detection of Multiple Pesticides Utilizing Gold Nanoparticles and Surface-Enhanced Raman Spectroscopy. *J Agric Food Chem* 67 (46), 12642–12651. <https://doi.org/10.1021/acs.jafc.9b01544>.
- [32] Yaseen, T., Pu, H., Sun, D.-W., 2019. Fabrication of silver-coated gold nanoparticles to simultaneously detect multi-class insecticide residues in peach with SERS technique. *Talanta* 196, 537–545. <https://doi.org/10.1016/j.talanta.2018.12.030>.
- [33] Chen, X., Wang, D., Li, J., Xu, T., Lai, K., Ding, Q., Lin, H., Sun, L., Lin, M., 2020. A spectroscopic approach to detect and quantify phosmet residues in Oolong tea by surface-enhanced Raman scattering and silver nanoparticle substrate. *Food Chem* 312, 126016. <https://doi.org/10.1016/j.foodchem.2019.126016>.
- [34] Sokolova, M., Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. *Inf Process Manag* 45 (4), 427–437. <https://doi.org/10.1016/j.ipm.2009.03.002>.
- [35] A. Botchkarev, 2018. Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology. arXiv 1809, 03006. <https://doi.org/https://doi.org/10.28945/4184>.
- [36] Waskom, M.L., 2021. Seaborn: statistical data visualization. *J Open Source Softw* 6 (60), 3021. <https://doi.org/10.21105/joss.03021>.
- [37] Jayalakshmi, T., Santhakumaran, A., 2011. Statistical normalization and back propagation for classification. *Int J Comput Theory Eng* 3 (1), 1793–8201. <https://doi.org/10.7763/IJCTE.2011.V3.288>.
- [38] Krishna, H., 2017. *Digital Signal Processing Algorithms: Number Theory, Convolution, Fast Fourier Transforms, and Applications*. Routledge,.