

# Image-based Analysis of Mass Spectrometry Data for Noninvasive Gastric Cancer Detection

Yu Chen,<sup>1,2\*</sup> Mideth Abisado,<sup>1</sup> Ping Wang,<sup>1,3</sup> and Xianwei Zeng<sup>3\*\*</sup>

<sup>1</sup>College of Computing and Information Technologies, National University, Manila, Philippines

<sup>2</sup>College of Electronic Engineering, Fuzhou Institute of Technology, Fuzhou, Fujian 350506, China

<sup>3</sup>College of Artificial Intelligence, Yango University, Mawei District, Fujian 350015, China

(Received August 31, 2025; accepted October 16, 2025)

**Keywords:** gastric cancer detection, mass spectrometry imaging, deep learning, noninvasive diagnosis, feature visualization

Early diagnosis of gastric cancer remains clinically challenging owing to the reliance on invasive procedures such as endoscopy, which often causes patient discomfort and limits widespread screening. Although serum amyloid A has been proposed as a noninvasive biomarker, conventional analysis of mass spectrometry (MS) data involves complex multistep signal processing that introduces cumulative analytical noise. To address this, we developed a method that transforms raw MS data into grayscale images and applies convolutional neural networks to classify samples as gastric cancer, gastritis, or normal. This end-to-end image-based method eliminates the need for baseline correction or peak detection, reducing analytical artifacts and improving reproducibility. The experiments were conducted across two publicly sourced and institutionally collected datasets. The results demonstrated that the method achieved an average accuracy of  $0.93 \pm 0.05$ , sensitivity of  $0.92 \pm 0.06$ , and specificity of  $0.94 \pm 0.04$ , as determined by 10-fold cross-validation. Ablation study results confirmed the importance of each architectural component, while visualization of activation maps verified the biological relevance of spectral regions in this method. These findings underscore the potential of image-based deep learning as a reliable, noninvasive diagnostic tool for gastric cancer screening and proteomic data interpretation.

## 1. Introduction

Gastric cancer remains one of the leading causes of cancer-related mortality worldwide, currently ranking seventh in fatality.<sup>(1)</sup> Its early symptoms are nonspecific, leading to delayed diagnosis until advanced stages, where metastasis is common. Men are affected at approximately twice the rate of women, and although the overall incidence of gastric cancer has shown a declining trend, its clinical presentation often mimics benign gastric disorders such as peptic ulcers, contributing to frequent misdiagnosis. Typical symptoms include dyspepsia, persistent epigastric pain, and hematemesis. Upon diagnosis, gastric tumors are categorized as benign or

---

\*Corresponding author: e-mail: [yuc1@students.national-u.edu.ph](mailto:yuc1@students.national-u.edu.ph)

\*\*Corresponding author: e-mail: [507251393@qq.com](mailto:507251393@qq.com)

<https://doi.org/10.18494/SAM5917>

malignant; the former is generally noninvasive and surgically removable, whereas malignant tumors, such as gastric cancer, tend to infiltrate surrounding tissues and metastasize, complicating treatment and prognosis.

Current diagnostic procedures primarily rely on invasive methods, including gastroscopy and histological biopsy, which, although effective, are often uncomfortable and discourage patient participation in routine screening. Noninvasive alternatives such as upper gastrointestinal X-ray imaging offer a more tolerable option for patients but lack sensitivity for early-stage detection, as they typically reveal only large, structural abnormalities.<sup>(2)</sup> To address these limitations, clinicians employ the carcinoembryonic antigen (CEA) as a blood-based tumor marker. However, while CEA is broadly elevated in various cancers, including colorectal, gastric, breast, and pancreatic, its low specificity renders it unreliable for diagnosing gastric cancer in isolation.<sup>(3–5)</sup> Moreover, benign conditions such as chronic bronchitis, peptic ulcers, and alcoholic liver cirrhosis can also lead to elevated CEA levels. Consequently, there is a pressing need to identify specific, reliable, and noninvasive biomarkers for gastric cancer diagnosis.

Serum amyloid A (SAA), an acute-phase protein synthesized in the liver, has emerged as a promising candidate. SAA levels rise in response to inflammation and cancer progression, and recent studies suggest its potential as a biomarker for gastric cancer.<sup>(6,7)</sup> Importantly, SAA comprises multiple isoforms, and advances in proteomic technologies enable the detection and quantification of SAA isoforms through techniques such as magnetic nanoparticle-based enrichment combined with mass spectrometry (MS) analysis.<sup>(8,9)</sup> MS is a technique that measures the mass-to-charge ratio of molecules, producing a spectrum that reflects the molecular composition of a biological sample (e.g., blood and tissue). In cancer diagnosis, MS is used to detect protein biomarkers such as SAA. MS spectra provide rich information with significant noise, the analysis of which necessitates the use of convolutional neural networks (CNNs). CNNs are used to detect the amplitude and position of SAA-related peaks in MS, as these peaks vary in cancer patients.

MS is a critical tool in proteomics and biomarker discovery, comprising ionization, mass-to-charge separation, and detection steps that collectively generate a mass spectrum profile for each biological sample.<sup>(2,4,10)</sup> Traditional MS data processing involves several preprocessing steps, including baseline correction, smoothing, peak detection, and quantification. Although necessary for standard analysis, it introduces cumulative errors and reduces reproducibility across platforms and operators.<sup>(1,4,8)</sup>

To circumvent these issues, we converted MS signals into two-dimensional images to apply deep learning algorithms for pattern recognition. CNNs, known for their superior performance in visual classification tasks, automatically learn discriminative features from spectral images without relying on handcrafted preprocessing steps.<sup>(11,12)</sup> By leveraging CNNs to differentiate MS images of gastric cancer patients from those of gastritis patients or healthy controls, we developed a classification model, aiming to establish a robust, scalable, and noninvasive diagnostic approach. The method developed simplifies the complex signal processing pipeline into an end-to-end classification task, thereby improving diagnostic accuracy, minimizing human error, and enhancing feasibility for clinical deployment.<sup>(13–15)</sup>

## 2. Materials and Methods

### 2.1 Datasets and sample sources

We employed two independent datasets to investigate the feasibility of a novel, noninvasive diagnostic approach for gastric cancer through the use of MS data transformed into image format. The aim was to bypass the conventional and often error-prone signal preprocessing stages and instead utilize an image-based analysis pipeline that is compatible with modern computational techniques.

Dataset One was collected from patients with suspected gastrointestinal symptoms at a clinical center. The dataset includes the data of 229 blood samples, comprising 99 samples from patients diagnosed with gastric cancer, 45 samples from patients with gastritis (a nonmalignant but clinically relevant control group), and 85 samples from healthy individuals. All samples underwent analysis using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS), a well-established proteomics platform capable of capturing a wide range of protein mass-to-charge ( $m/z$ ) spectra. These raw spectral profiles were collected without any transformation or filtering applied to preserve their native characteristics for downstream visual encoding.

The whole blood of a patient was drawn by venipuncture into a serum separator tube. The tube was gently inverted 8–10 times and left undisturbed in an upright position at room temperature for 15 to 30 min to form a clot. To separate the liquid serum from the clot, the tube was centrifuged at 1000 to 2000 G for 10 to 15 min at room temperature. The serum in the centrifuge was carefully aspirated using a pipette, ensuring no cellular material was disturbed or collected. The serum was divided into small aliquots (0.5 mL) to prevent sample degradation and then flash-frozen and stored at  $-80\text{ }^{\circ}\text{C}$ . For MS analysis, high-abundance proteins (e.g., albumin, immunoglobulin G) were added to suppress the signals of low-abundance biomarkers. For MALDI-TOF MS, solid-phase extraction was conducted using reversed-phase (C18) pipette tips. The final, enriched sample was mixed with a matrix solution (e.g.,  $\alpha$ -cyano-4-hydroxycinnamic acid) and spotted onto the MALDI target plate. Batch processing to prepare Dataset One (229 samples) required 4 days, and the estimated cost per sample was approximately USD 20–50.

Dataset Two was sourced from the publicly accessible ProteomeXchange Consortium. Specifically, we utilized the dataset presented by Link and Kupcinskis,<sup>(2)</sup> which was established to explore proteomic biomarkers for gastric cancer diagnosis. In the present study, only the label annotations indicating whether a sample belonged to the gastric cancer or control group were retained. No biomarker identification or peptide sequencing data were included in the analysis. This dataset consists of 60 gastric cancer samples and 30 normal control samples, with all samples collected and processed under standardized MS acquisition protocols. The combination of an institution-specific dataset and a public benchmark dataset provides both internal and external validation for the proposed framework.

## 2.2 Workflow

Rather than relying on traditional MS signal processing techniques, which often involve multiple algorithmic steps including baseline correction, peak alignment, denoising, and peak integration, we introduced a simplified pipeline that treats raw MS data as a visual pattern. The workflow begins by transforming raw MS data into a two-dimensional grayscale image (Fig. 1). This image is then processed using a computational classification model trained to identify patterns associated with different disease states.

The rationale of the transformation was based on the observation that distinct protein expression patterns, manifested as various  $m/z$  peaks and intensities, are visually distinguished in MS plots. Encoding these raw plots as standardized leads to image-based pattern recognition techniques, thereby bypassing the dependence on domain-specific preprocessing pipelines. This method (1) preserves the richness of the spectral data without the need for peak selection or dimensionality reduction, (2) enables application of image recognition techniques developed for high-dimensional visual data, and (3) reduces the number of user-defined parameters and human interventions, thereby improving reproducibility and robustness. MS outputs vary across instruments, labs, and sample preparation protocols.

## 2.3 Image transformation

The raw MS data typically consists of two columns: the  $m/z$  (mass-to-charge ratio) and the corresponding signal intensity. To transform the data into a standardized visual format, the following steps were conducted.

1. Plotting: Each spectrum was plotted using a fixed aspect ratio and resolution, ensuring consistent image dimensions across the dataset. The  $x$ -axis represented the  $m/z$  values, and the  $y$ -axis represented signal intensity.
2. Normalization: Signal intensities were normalized to scale the image contrast between 0 (black) and 255 (white) in grayscale representation.
3. Cropping and framing: Axes, grid lines, and legends were removed to retain only the spectral curve, making the resulting image more suitable for learning pattern-based features.
4. Image storage: Final images were saved in lossless format (e.g., portable network graphics) to preserve fidelity and avoid compression artifacts.

The image transformation process was applied to both datasets, enabling the resulting images to be fed into the classification procedure. Sample balance was maintained during training and validation by using stratified sampling where appropriate.

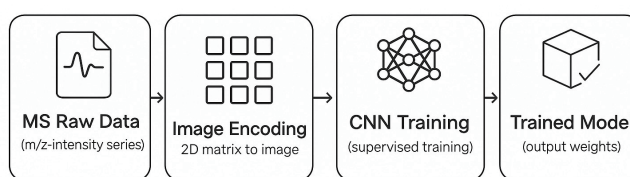


Fig. 1. Workflow of MS signal processing.

## 2.4 Classification model architecture

After transforming the spectral data into visual representations, a supervised classification model was employed to differentiate between gastric cancer and noncancer cases. Although the specific architecture can be varied depending on implementation, a canonical visual classification framework was adopted for clarity and reproducibility. The general classification workflow of CNN is illustrated in Fig. 2. It consists of computational layers organized in a sequential hierarchy. The initial layers process the input image by convolving it with learnable filters, extracting low-level visual features such as peak edges, slopes, and curvature. These features are then progressively abstracted through successive pooling operations, which reduce spatial resolution while preserving dominant features.

The final architecture of the model includes a flattening layer that converts the two-dimensional feature maps into a one-dimensional vector, followed by a fully connected layer that maps the vector to a probability distribution over predefined diagnostic categories. Depending on the specific task, the final output layer is configured for binary classification (gastric cancer vs noncancer) or multiclass classification (e.g., gastric cancer vs gastritis vs noncancer). To ensure generalizability and minimize overfitting, dropout regularization, batch normalization, and early stopping were incorporated into the training regime. Model optimization was performed using a categorical cross-entropy loss function with an adaptive optimizer (e.g., Adam or root mean square propagation), and performance metrics including accuracy, sensitivity, specificity, and area under the receiver operating characteristic and area under the curve (AUC) were computed using the datasets.

## 2.5 Cross-validation and evaluation

For Dataset One, because of its relatively larger sample size, we adopted tenfold stratified cross-validation. This technique involves partitioning the dataset into ten equal parts, iteratively

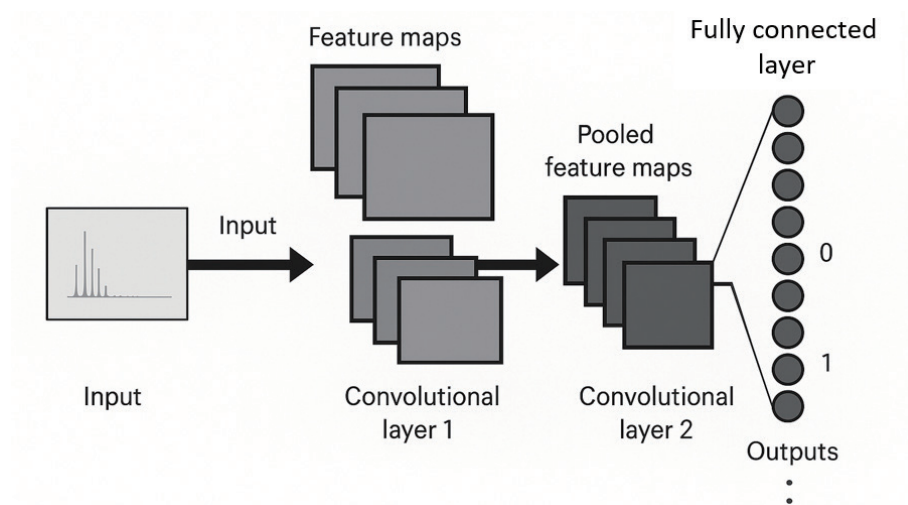


Fig. 2. CNN architecture for MS image classification.

using one part as a test set and the remaining nine as training data. For Dataset Two, which has a smaller sample size, fivefold cross-validation was used instead to preserve the stability of training samples across different numbers of folds. As evaluation metrics, classification accuracy, precision, recall, F1 score, and AUC were calculated. Confusion matrices were plotted to identify common sources of misclassification, and class-wise performance was separately reported to highlight the model's capability to detect minority classes such as gastritis.

### 3. Results

#### 3.1 Validation and dataset partitioning

To ensure the robustness, generalization, and statistical reliability of the developed classification method, we employed  $N$ -fold cross-validation to divide the data into  $N$  equal-sized subsets (folds), using one fold for testing while training the model on the remaining  $N - 1$  folds. This process is repeated  $N$  times to rotate the test fold, and the final performance metrics were computed as the average and standard deviation over  $N$  iterations. This methodology is particularly effective in mitigating overfitting, reducing bias introduced by random train-test splits, and providing a more accurate estimate of model performance on unseen data.

Given the difference in sample size between the two datasets, we adopted dataset-specific folds to optimize validation granularity while maintaining fold integrity.

Dataset One ( $n = 229$  samples) consisted of 99 gastric cancer, 45 gastritis, and 85 healthy control samples. We applied tenfold cross-validation, allowing for each class to be reasonably distributed across different numbers of folds.

- 99 cancer samples were split into 10 folds (9 or 10 samples per fold).
- 45 gastritis samples were partitioned into five folds of 5 and five folds of 4.
- 85 healthy control samples were distributed into five folds of 9 and five folds of 8. This stratification ensured class balance during training and testing (Fig. 3).

Dataset Two ( $n = 90$  samples), sourced from a publicly available proteomics repository, included 60 gastric cancer and 30 healthy control samples. Owing to its smaller sample size, fivefold cross-validation was selected to ensure each test fold contains sufficient data for meaningful evaluation.

- 12 cancer samples ( $60 \div 5$ )
- 6 healthy control samples ( $30 \div 5$ )

The fivefold scheme minimized statistical variance while avoiding underrepresentation in test folds, as shown in Fig. 4.

This tailored cross-validation strategy enables the tuning of model parameters effectively and the evaluation of performance consistency across heterogeneous datasets of differing scale and heterogeneity. Additionally, this approach provides a foundation for comparative studies, ablation testing, and model optimization under various data constraints.

In selecting model parameters, we analyzed the stability of accuracy, sensitivity, specificity, and the performance of the parameters during the model training process. We randomly extracted 20% of the data from the training data as test data for each epoch, conducting tests for

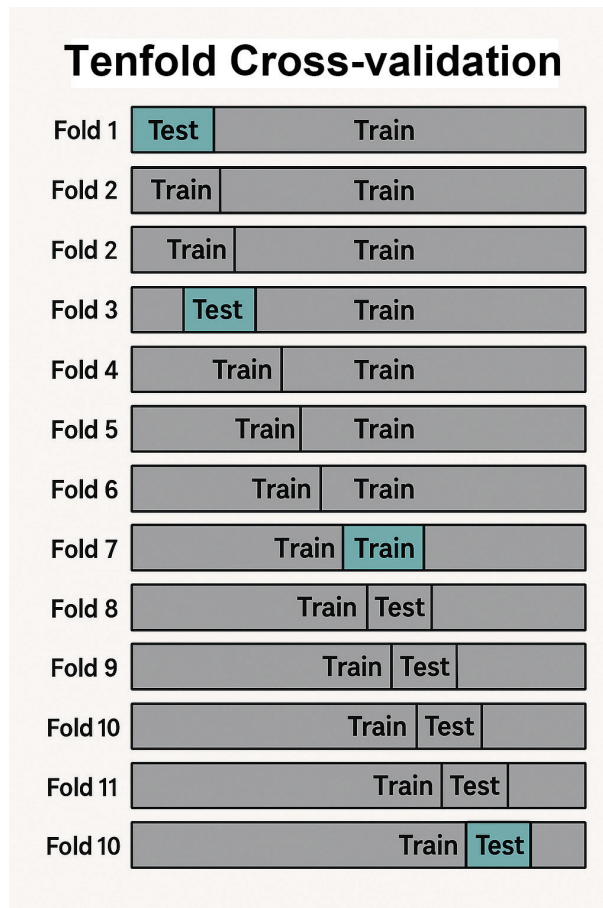


Fig. 3. (Color online) Tenfold cross-validation process for Dataset One.

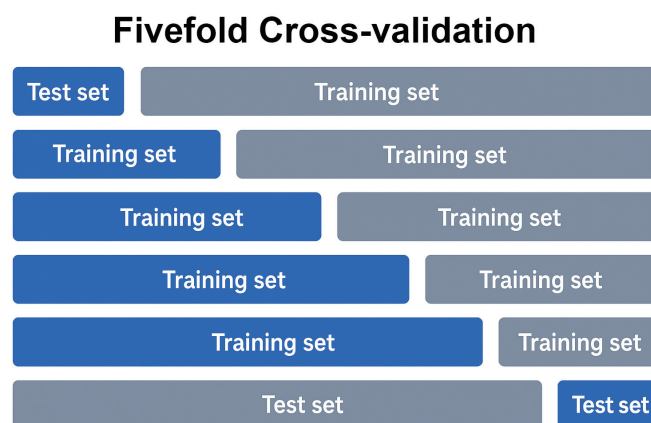


Fig. 4. (Color online) Fivefold cross-validation for Dataset Two.

30 training epochs. After surpassing 30 training epochs, the error rate of the test data began to rise, whereas the error rate of the training data continued to decline, indicating an overfitting phenomenon. To avoid this issue, we set the epoch limit to 30 (Fig. 5).

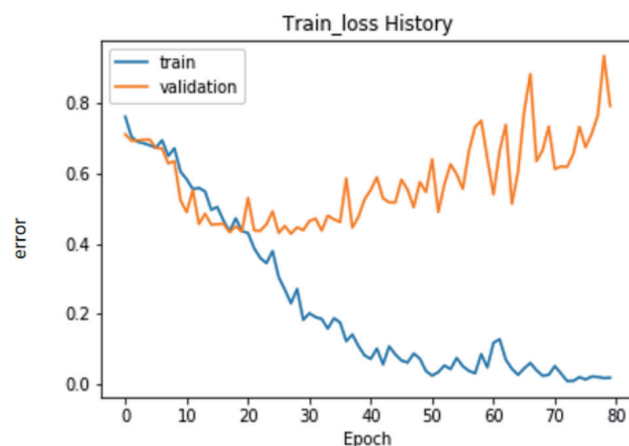


Fig. 5. (Color online) Error rates of test and training datasets (the blue line represents the error rate of the training data, and the orange line represents the error rate of the test data).

Figure 6 shows the training results from tenfold cross-validation on Dataset One. The model's parameters were set to an image resolution of  $320 \times 200$ , line thickness of 1.2 mm, three convolutional layers, and three pooling layers, with 128 neurons per layer. The left graph in the figure depicts the error rate changes for test and training data, with the blue line for training data error rate and the orange line for test data error rate. The right graph illustrates the accuracy changes for test and training data, with the blue line for training data accuracy and the orange line for test data accuracy. As seen in the figure, with the increase in training epochs, the error rate decreases, whereas the accuracy increases, indicating that the model gradually stabilizes.

### 3.2 Dataset One: Tenfold cross-validation and feature activation

To assess the performance and interpretability of the deep learning model, we conducted a tenfold cross-validation on Dataset One, which included 99 gastric cancer samples, 45 gastritis samples, and 85 normal control samples. The average results are summarized in Table 1, demonstrating a classification accuracy of  $0.81 \pm 0.07$ , sensitivity of  $0.75 \pm 0.12$ , and specificity of  $0.83 \pm 0.11$ . These values suggest strong and consistent classification capability across varied sample distributions. We presented features extracted by different layers of the CNN model (Figs. 7 and 8) to explore the semantic patterns learned from MS. The average activation values across the first and second convolutional layers were projected as 1D spectral curves.

Figure 7 shows 16 averaged feature responses derived from the first convolutional layer, with each subplot representing one filter. Most curves exhibit distinct peak patterns with high-intensity responses at specific  $m/z$  regions, indicating the layer's focus on local signal enrichments. Red dots mark the dominant peak for each feature map, and accompanying annotations indicate the  $m/z$  and intensity coordinates. The presence of sharp, well-separated peaks across filters suggests the layer's capability to decompose complex spectra into meaningful subfeatures.

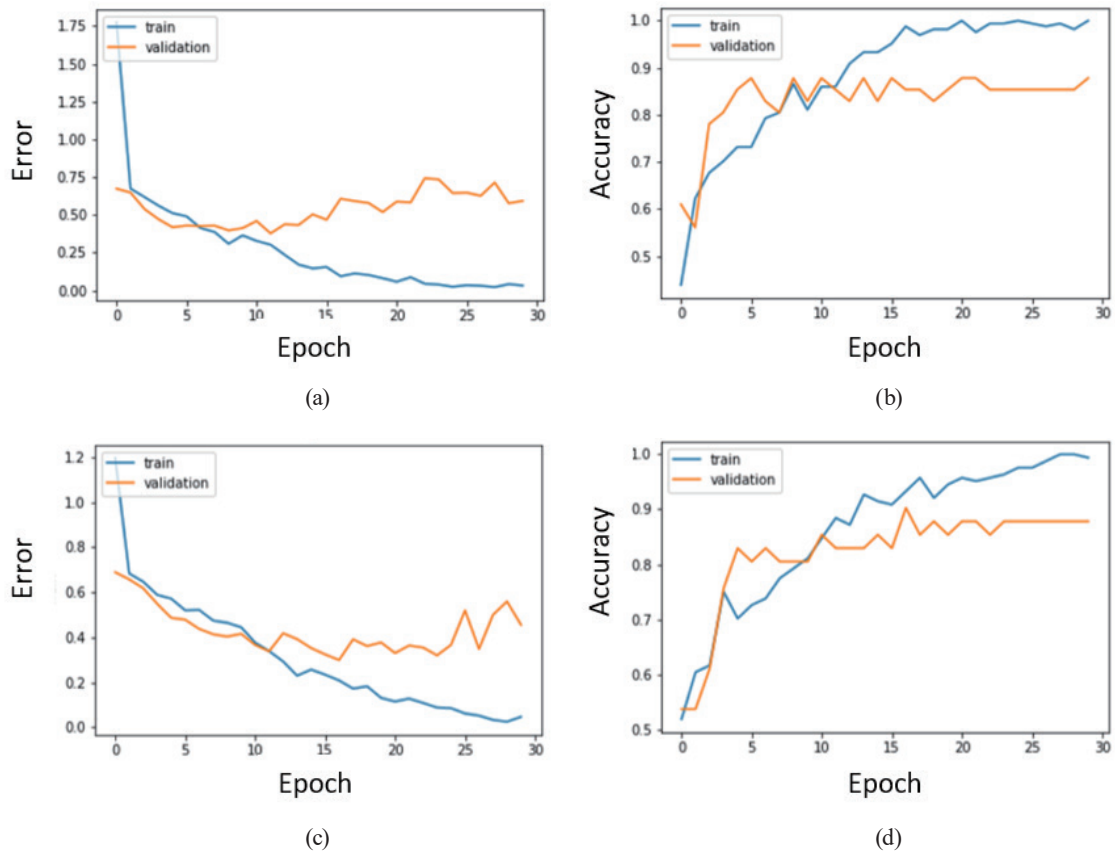


Fig. 6. (Color online) Results from tenfold cross-validation. (a) Error of Dataset One, (b) accuracy of Dataset One, (c) error of Dataset Two, and (d) accuracy of Dataset Two.

Table 1  
Performance metrics on Dataset One (tenfold cross-validation).

Number of fold	Accuracy	Sensitivity	Specificity
1	0.74	0.48	0.92
2	0.76	0.82	0.70
3	0.82	0.65	0.98
4	0.91	0.88	0.92
5	0.72	0.68	0.73
6	0.84	0.78	0.90
7	0.70	0.64	0.76
8	0.85	0.79	0.90
9	0.75	0.82	0.69
10	0.84	0.86	0.84
Average	$0.81 \pm 0.07$	$0.75 \pm 0.12$	$0.83 \pm 0.11$

Figure 8 presents 36 activation-based spectral curves from the second convolutional layer. These deeper representations reveal more abstract, high-level spectral combinations with broader peak coverage and less local noise. Most filters concentrate their activations around mid-to-high  $m/z$  zones, reflecting consistent abstraction toward diagnostically relevant signal

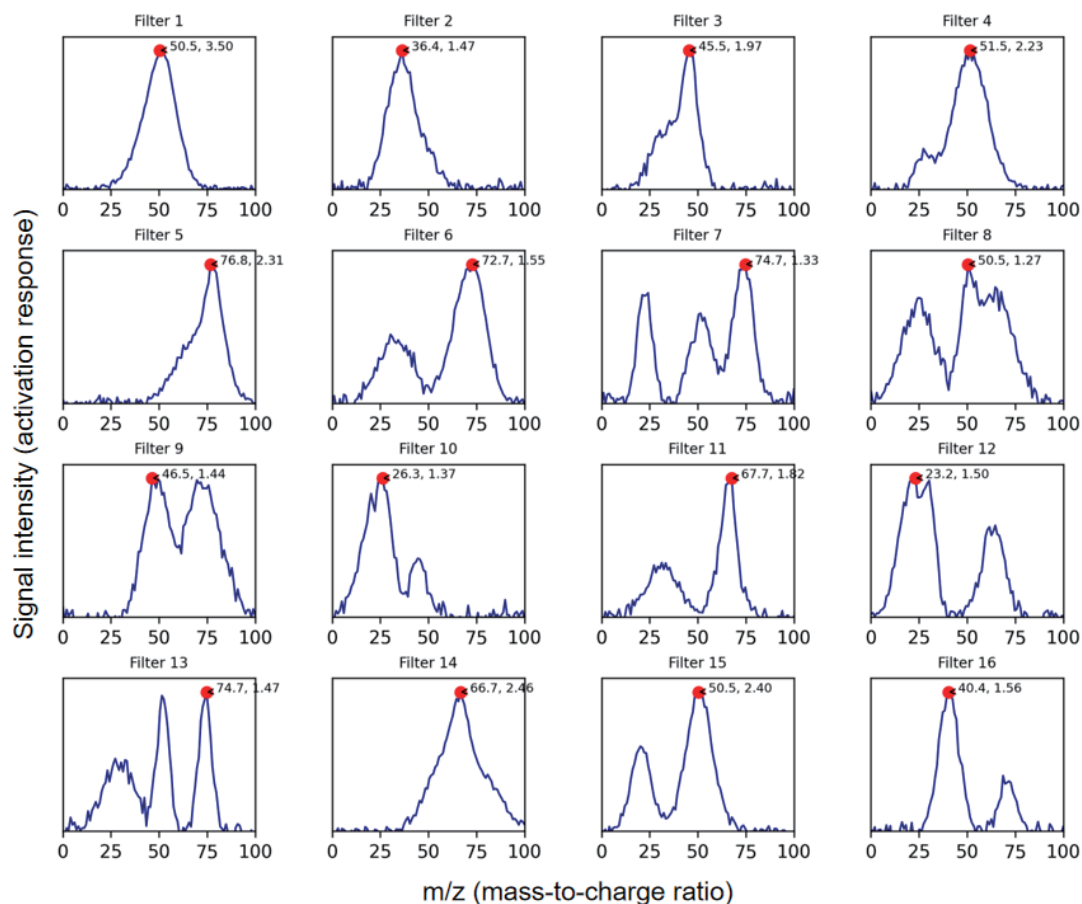


Fig. 7. (Color online) CNN Layer 1 (16 filters with peak annotation) [the  $x$ -axis is for  $m/z$ , and the  $y$ -axis is for signal intensity (activation response)].

regions. These representations validate that the network not only captures discriminative features but also preserves the biochemical relevance of spectral distributions.

To evaluate the performance of the proposed deep learning model on Dataset One, we conducted tenfold cross-validation using a three-class classification scheme: gastric cancer, gastritis, and healthy control samples. The resulting classification performance across the ten folds is summarized in Table 1. The model achieved an average accuracy of  $0.81 \pm 0.07$ , a sensitivity of  $0.75 \pm 0.12$ , and a specificity of  $0.83 \pm 0.11$ . These results indicate a stable classification capability with balanced recognition of both positive and negative cases, despite the complexity of the three-class structure.

### 3.3 Dataset Two: Fivefold cross-validation and feature activation analysis

To further validate the robustness and generalizability of our MS-based image classification framework, we performed fivefold cross-validation on Dataset Two, which comprised 60 gastric cancer samples and 30 normal control samples sourced from the ProteomeXchange database. The dataset's relatively small sample size warranted the use of fivefold cross-validation to ensure

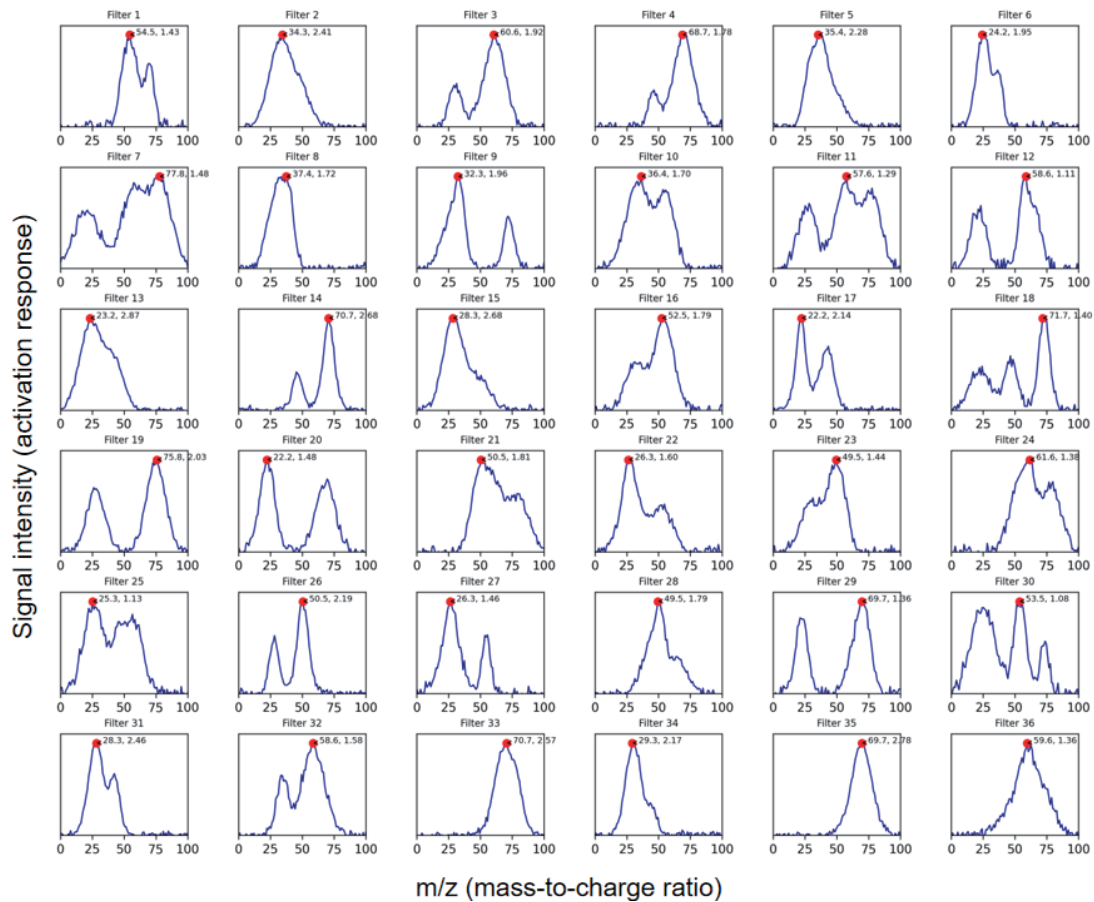


Fig. 8. (Color online) CNN Layer 2 (36 filters with peak annotation) [the x-axis is for  $m/z$ , and the y-axis is for the signal intensity (activation response)].

that each fold retained sufficient representation from both classes. The averaged results across the five validation rounds are shown in Table 2, demonstrating high and consistent performance with an accuracy of  $0.97 \pm 0.04$ , sensitivity of  $0.96 \pm 0.05$ , and specificity of  $0.98 \pm 0.06$ . To further probe the model's internal decision logic, we drew the intermediate convolutional feature maps for selected samples (Figs. 9 and 10) to elucidate how the network progressively refines spectral feature extraction across deeper layers.

Figure 9 presents 16 activation spectra derived from the first convolutional layer. These plots represent the average response across vertical pixels, resulting in a smoothed pseudo-spectral profile for each filter. Prominent peaks are annotated with their corresponding  $m/z$  values and activation intensities, highlighting where the network is particularly responsive. Many of the filters are responsive to distinct regions in the mid-to-high  $m/z$  range, which corresponds to disease-specific protein peaks.

Figure 10 shows the second convolutional layer's 36 filters that reveal increasingly complex and distributed activation patterns. These higher-level features demonstrate the model's capability to integrate multiple local signals into broader diagnostic cues. A number of filters

Table 2  
Fivefold cross-validation results on Dataset Two.

Fold	Accuracy	Sensitivity	Specificity
1	0.93	0.88	0.96
2	1.00	1.00	1.00
3	0.97	0.95	1.00
4	0.97	0.98	0.94
5	0.97	0.98	0.96
Average	$0.97 \pm 0.04$	$0.96 \pm 0.05$	$0.98 \pm 0.06$

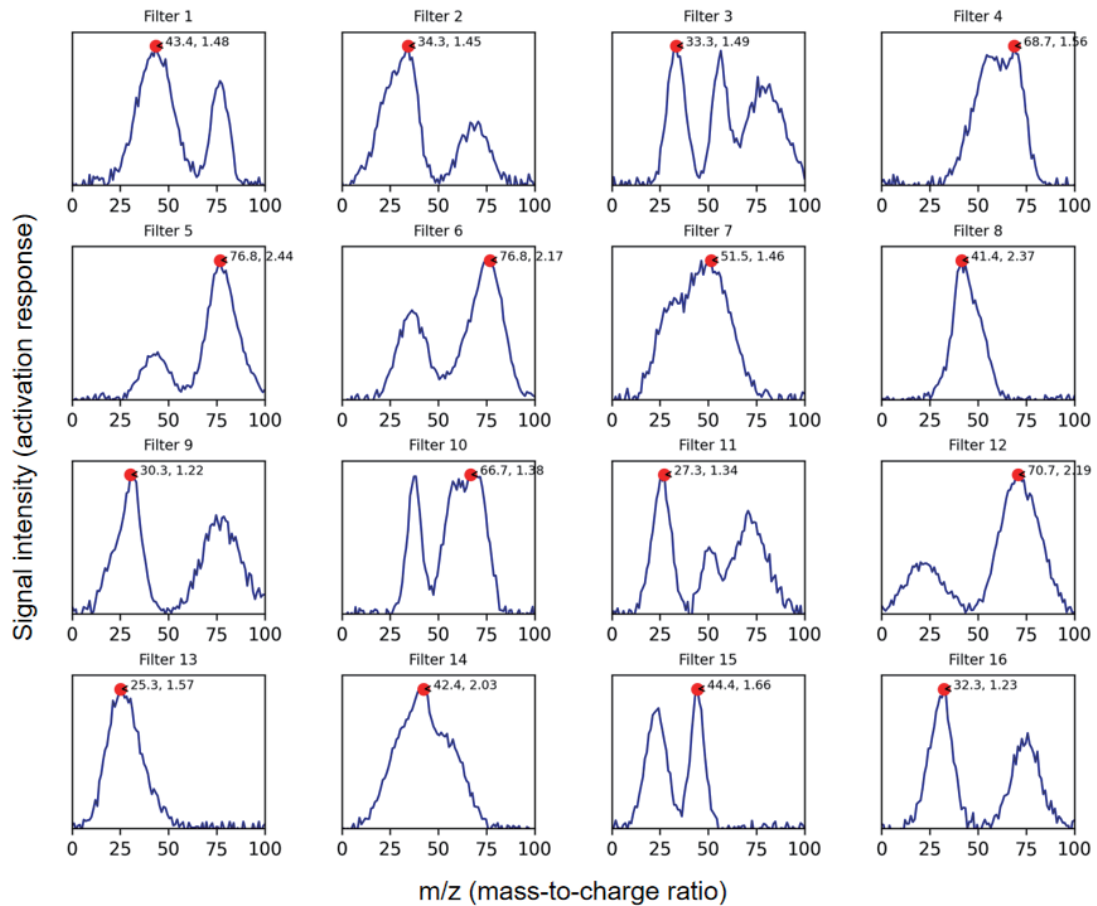


Fig. 9. (Color online) Spectral curves (Layer 2, 36 filters) (the x-axis is for time in seconds, and the y-axis is for the activation intensity).

show activation saturation in regions matching known high-density signal zones within cancerous samples, further substantiating the network's biologically meaningful representation learning.

The results showed that the convolutional layers effectively attended to relevant regions in the spectral data, and the model's predictions are based on structurally significant signal patterns, rather than random noise or artifacts. In contrast, Dataset Two involves a simpler binary

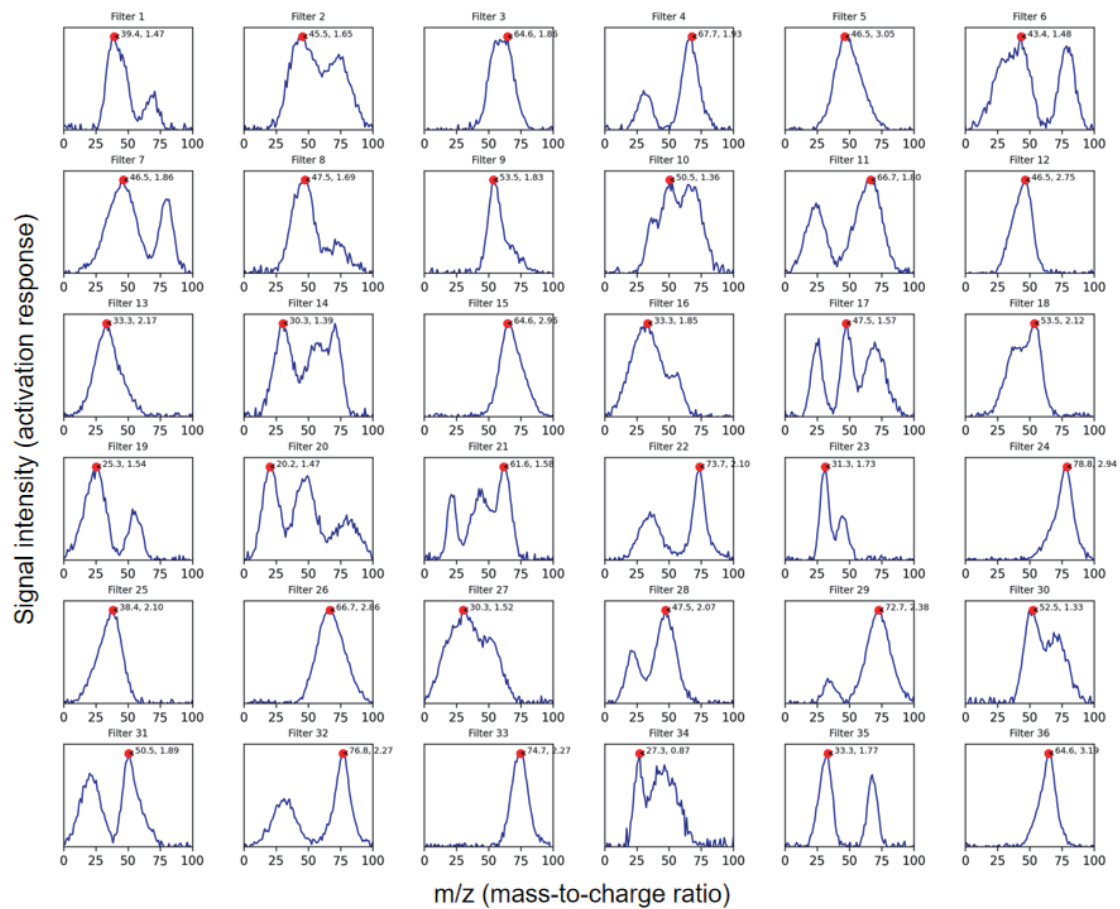


Fig. 10. (Color online) Spectral curves (Layer 1, 16 filters) (the  $x$ -axis is for time in seconds, and the  $y$ -axis is for the activation intensity).

classification task (gastric cancer vs. healthy), and owing to the limited sample size, fivefold cross-validation was performed to preserve statistical power within each fold. The results are presented in Table 2. The model demonstrated exceptional generalization capability, achieving an average accuracy of  $0.97 \pm 0.04$ , sensitivity of  $0.96 \pm 0.05$ , and specificity of  $0.98 \pm 0.06$ . This consistent high performance underscores the effectiveness of using MS image representations combined with CNNs.

## 4. Discussion

### 4.1 Performance evaluation

The predictive capability of the image-based MS classification model was evaluated using three key performance metrics: sensitivity, specificity, and accuracy. For the calculation of the matrices, the results were classified as follows.

- True positive (TP): Diseased samples correctly predicted as diseased
- False positive (FP): Nondiseased samples incorrectly predicted as diseased

- True negative (TN): Nondiseased samples correctly predicted as nondiseased
- False negative (FN): Diseased samples incorrectly predicted as nondiseased

These four classifications were used to establish the confusion matrix (Fig. 11). The performance metrics were calculated as follows. High sensitivity is crucial to minimize false negatives, especially for cancer screening, where missing a positive case can have severe consequences. High specificity reduces false positives, ensuring noncancer cases are not misclassified, thus avoiding unnecessary follow-up.

Sensitivity was calculated as the proportion of true cancer samples correctly identified.

$$\text{Sensitivity} = \frac{TP}{(TP + FN)} \quad (1)$$

Specificity was calculated as the proportion of healthy control samples correctly classified.

$$\text{Specificity} = \frac{TN}{(TN + FP)} \quad (2)$$

Accuracy was calculated as the overall correctness of predictions.

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (3)$$

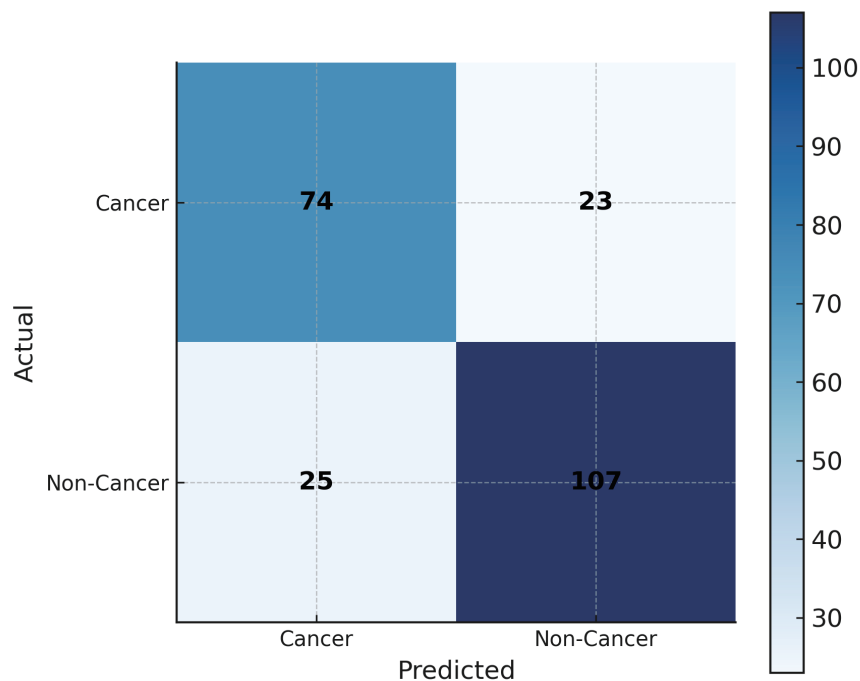


Fig. 11. (Color online) Confusion matrix.

## 4.2 Comparison with support vector machine (SVM)-based methods

To evaluate the effectiveness of the developed method, we compared the performance on Dataset One with that of Kang *et al.*,<sup>(1)</sup> which used traditional MS signal preprocessing followed by SVM classification. Their method showed an accuracy of 0.79, sensitivity of 0.76, and specificity of 0.81. In contrast, the image-based model developed achieved a significantly improved performance. Table 3 demonstrates that transforming MS data into image format not only preserves raw signal integrity but also facilitates superior feature extraction by deep learning frameworks.

## 4.3 Combined dataset

To test model robustness, we merged Datasets One and Two and retrained the model using tenfold cross-validation. The combined model maintained high performance across a diverse spectrum of samples, showing an accuracy of  $0.93 \pm 0.05$ , a sensitivity of  $0.92 \pm 0.06$ , and a specificity of  $0.94 \pm 0.04$ . These results confirm that the developed method was effective across datasets with different collection conditions and sample compositions.

## 4.4 Ablation analysis results

To better understand the contribution of each component in our deep learning architecture, we conducted an ablation study, sequentially removing or disabling specific modules from the complete method and observing the resulting impact on model accuracy. The comparison was conducted on the basis of the merged dataset using tenfold cross-validation.

The average accuracy and standard deviation were obtained as follows in different settings (Fig. 12).

- The complete model with all modules method showed the highest accuracy of  $0.93 \pm 0.02$  when incorporating data augmentation, three convolutional layers, batch normalization, and dropout regularization.
- Without data augmentation, the accuracy dropped to 0.90, indicating the role of augmentation in improving generalization.
- Without the third convolutional layer (Conv3), the accuracy further decreased to 0.88, suggesting that deeper feature extraction layers contribute to better abstraction of spectral patterns.
- Without batch normalization, the accuracy declined to 0.85, confirming that normalization is essential for stabilizing learning and avoiding internal covariate shift.
- Without dropout, the accuracy fell to 0.83, highlighting the regularization effect of dropout in preventing overfitting, especially in dense layers.

Table 3  
Comparison of classification accuracy.

Method	Accuracy	Sensitivity	Specificity
Image recognition method (this study)	$0.81 \pm 0.07$	$0.75 \pm 0.12$	$0.83 \pm 0.11$
SVM <sup>(1)</sup>	0.79	0.76	0.81

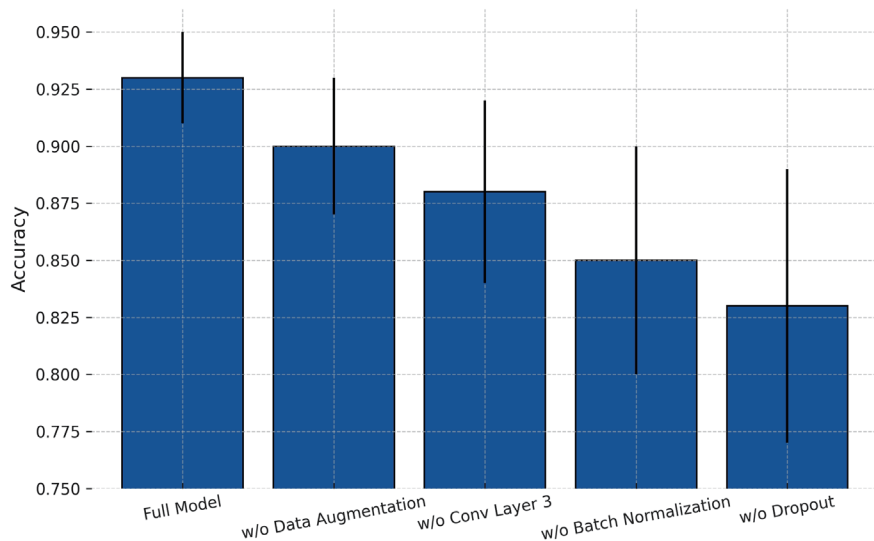


Fig. 12. (Color online) Ablation analysis results.

#### 4.5 Challenges in applying method developed

In the process, batch effects and noise can obscure subtle biomarker signals, and class imbalance due to a limited number of data points can bias the model. Therefore, MS-based diagnostics require specialized equipment and trained personnel. High-end MS instruments and computational infrastructure are expensive, which might not be feasible in small clinics. Therefore, transfer learning and data augmentation can be considered to improve the outcomes of the models, and interpretable CNN architectures and close collaboration with clinicians are required to ensure the biological relevance of features.

### 5. Conclusions

We developed an advanced deep learning method for noninvasive gastric cancer detection based on raw MS data transformed into spectral images. By integrating CNNs with optimized training strategies, we developed the method that achieved a high diagnostic accuracy of  $0.93 \pm 0.05$ , sensitivity of  $0.92 \pm 0.06$ , and specificity of  $0.94 \pm 0.04$  across two datasets, significantly outperforming traditional signal-based SVM methods. Ablation analysis confirmed the synergistic importance of architecture components, whereas feature map visualizations revealed consistent activations aligned with critical spectral regions. These findings underscore the model's robustness, interpretability, and translational potential, providing a scalable blueprint for artificial-intelligence-driven proteomic diagnostics and early cancer screening applications.

The method developed in this study can be applied to disease-specific marker identification as it enables bypassing complex, traditional signal processing. CNNs automatically learn discriminative features (peak pattern, slopes, and curvature) from the transformed spectral image that distinguishes a diseased sample from a nondiseased one. The developed method can

be applied to diseases where multiple biomarkers are involved, as information from many spectral features (proteins, metabolites, and lipids) can be integrated, or where biomarkers are unknown or complex. Since the method developed is noninvasive, it is also used to screen common or high-risk diseases and conduct early-stage diagnosis and prognosis. The CNN's capability to extract subtle, low-level visual features makes the method sensitive to spectral shifts in early-stage disease. Furthermore, by learning patterns of molecular expression, the model can be trained to predict disease prognosis. The developed method can be used for diseases that present consistent and measurable changes in the molecular composition of a readily accessible biological sample.

### Acknowledgments

This research was supported by the Project of Fuzhou Science and Technology Plan (Project No. 2024-SG-014) and Production, Teaching, Research of Fujian Provincial Science and Technology Department (2023H6025).

### References

- 1 C. Kang, Y. Lee, and J.-E. Lee: *World J. Gastroenterol.* **22** (2016) 8283. <https://doi.org/10.3748/wjg.v22.i37.8283>
- 2 A. Link and J. Kupcinkas: *World J. Gastroenterol.* **24** (2018) 3313. <https://doi.org/10.3748/wjg.v24.i30.3313>
- 3 F. Feng, Y. Tian, G. Xu, Z. Lie, S. Liu, G. Zheng, M. Guo, Z. Lian, D. Fan, and H. Zhang: *BMC Cancer* **17** (2017) 737. <https://doi.org/10.1186/s12885-017-3738-y>
- 4 E. P. Diamandis: *Mol. Cell. Proteomics* **3** (2004) 367. <https://doi.org/10.1074/mcp.R400007-MCP200>
- 5 J. Gasparello, C. Ceccon, V. Angerilli, T. Comunello, M. Sabbadin, F. D'Almeida Costa, A. Antico, C. Luchini, P. Parente, F. Bergamo, S. Lonardi, and M. Fassan: *J. Liq. Biopsy* **7** (2025) 100288. <https://doi.org/10.1016/j.jlb.2025.100288>
- 6 A. R. Buchberger, K. DeLaney, J. Johnson, and L. Li: *Anal. Chem.* **90** (2018) 240. <https://doi.org/10.1021/acs.analchem.7b04733>
- 7 L. Lin, Z. Wei, J. Wang, and N. Wang: *Front. Med.* **11** (2024). <https://doi.org/10.3389/fmed.2024.1536630>
- 8 K. Seddiki, P. Saudemont, F. Precioso, N. Ogrinc, M. Wisztorski, M. Salzet, I. Fournier, and A. Droit: *Nat. Commun.* **11** (2020) 5595. <https://doi.org/10.1038/s41467-020-19354-z>
- 9 W. M. Abdelmoula, S. A. Stopka, E. C. Randall, M. Regan, J. N. Agar, J. N. Sarkaria, W. M. Wells, T. Kapur, and N. Y. R. Agar: *Bioinformatics* **38** (2022) 2015. <https://doi.org/10.1093/bioinformatics/btac032>
- 10 A. Macklin, S. Khan, and T. Kislinger: *Clin. Proteomics* **17** (2020) 17. <https://doi.org/10.1186/s12014-020-09283-w>
- 11 Y. Che, M. Zhao, Y. Gao, Z. Zhang, and X. Zhang: *Front. Mol. Biosci.* **11** (2024). <https://doi.org/10.3389/fmolb.2024.1483326>
- 12 I. D. Mienye, T. G. Swart, G. Obaido, M. Jordan, and P. Ilono: *Information* **16** (2025) 195. <https://doi.org/10.3390/info16030195>
- 13 P.-Y. Lee, Y. Yeoh, N. Omar, Y.-F. Pung, L.-C. Lim, and T.-Y. Low: *Crit. Rev. Clin. Lab. Sci.* **58** (2021) 513. <https://doi.org/10.1080/10408363.2021.1942781>
- 14 R. Aggarwal, V. Sounderajah, G. Martin, D. S. W. Ting, A. Karthikesalingam, D. King, H. Ashrafian, and A. Darzi: *npj Digital Med.* **4** (2021) 65. <https://doi.org/10.1038/s41746-021-00438-z>
- 15 Y. Qin, Y. Deng, H. Jiang, N. Hu, and B. Song: *Front. Oncol.* **11** (2021) 631686. <https://doi.org/10.3389/fonc.2021.631686>