

## Article

# MSIMG: A Density-Aware Multi-Channel Image Representation Method for Mass Spectrometry

Fengyi Zhang <sup>1</sup>, Boyong Gao <sup>1</sup>, Yinchu Wang <sup>2,3,4</sup>, Lin Guo <sup>2,3,4</sup>, Wei Zhang <sup>2,3,4</sup> and Xingchuang Xiong <sup>2,3,4,\*</sup><sup>1</sup> College of Information Engineering, China Jiliang University, Hangzhou 310018, China; fengyi.zhang@cjl.u.edu.cn (F.Z.)<sup>2</sup> National Institute of Metrology, Beijing 100029, China<sup>3</sup> Key Laboratory of Metrology Digitalization and Digital Metrology for State Market Regulation, State Administration for Market Regulation, Beijing 100029, China<sup>4</sup> National Metrology Data Center, Beijing 100029, China

\* Correspondence: xiongxch@nim.ac.cn

## Abstract

Extracting key features for phenotype classification from high-dimensional and complex mass spectrometry (MS) data presents a significant challenge. Conventional data representation methods, such as traditional peak lists or grid-based imaging strategies, are often hampered by information loss and compromised signal integrity, thereby limiting the performance of downstream deep learning models. To address this issue, we propose a novel data representation framework named MSIMG. Inspired by object detection in computer vision, MSIMG introduces a data-driven, “density-peak-centric” patch selection strategy. This strategy employs density map estimation and non-maximum suppression algorithms to locate the centers of signal-dense regions, which serve as anchors for dynamic, content-aware patch extraction. This process transforms raw mass spectrometry data into a multi-channel image representation with higher information fidelity. Extensive experiments conducted on two public clinical mass spectrometry datasets demonstrate that MSIMG significantly outperforms both the traditional peak list method and the grid-based MetImage approach. This study confirms that the MSIMG framework, through its content-aware patch selection, provides a more information-dense and discriminative data representation paradigm for deep learning models. Our findings highlight the decisive impact of data representation on model performance and successfully demonstrate the immense potential of applying computer vision strategies to analytical chemistry data, paving the way for the development of more robust and precise clinical diagnostic models.

**Keywords:** mass spectrometry; density-aware; multi-channel image representation; deep learning



Academic Editor: Tomasz Marcin Dymerski

Received: 17 September 2025

Revised: 8 October 2025

Accepted: 14 October 2025

Published: 15 October 2025

**Citation:** Zhang, F.; Gao, B.; Wang, Y.; Guo, L.; Zhang, W.; Xiong, X. MSIMG: A Density-Aware Multi-Channel Image Representation Method for Mass Spectrometry. *Sensors* **2025**, *25*, 6363. <https://doi.org/10.3390/s25206363>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Mass Spectrometry (MS), a core tool in analytical chemistry, plays a pivotal role in frontier life science fields such as metabolomics [1–4], proteomics [5–7], and disease diagnostics [8–10]. This technology detects thousands of molecules in biological samples with high sensitivity and resolution, resulting in MS data characterized by high dimensionality, complexity, and large volume [11]. A significant computational challenge in the field is how to efficiently and accurately extract key features associated with specific biological states, such as disease, from this massive and intricate dataset [12]. Consequently, the development of novel and effective computational methods for analyzing MS data is vital

not only for elucidating complex biological mechanisms but also for facilitating the early diagnosis of diseases and enabling personalized medicine.

Traditional MS data analysis workflows typically rely on a series of steps including peak detection, metabolite identification, and statistical analysis [13], but this process faces numerous inherent difficulties. For instance, incomplete metabolite annotations lead to the neglect of a large number of signals, algorithmic biases in peak detection and alignment can introduce errors, and unavoidable signal shifts during data acquisition severely affect research reproducibility. Concurrently, artificial intelligence technologies, particularly deep learning [14], have made groundbreaking advancements in fields like computer vision. Models such as Convolutional Neural Networks (CNNs) have demonstrated superhuman performance in tasks like image recognition [15]. Their powerful feature extraction and pattern recognition capabilities offer new approaches for solving complex data analysis problems in other domains. Therefore, integrating MS data with computer vision techniques represents a promising interdisciplinary research direction [16,17].

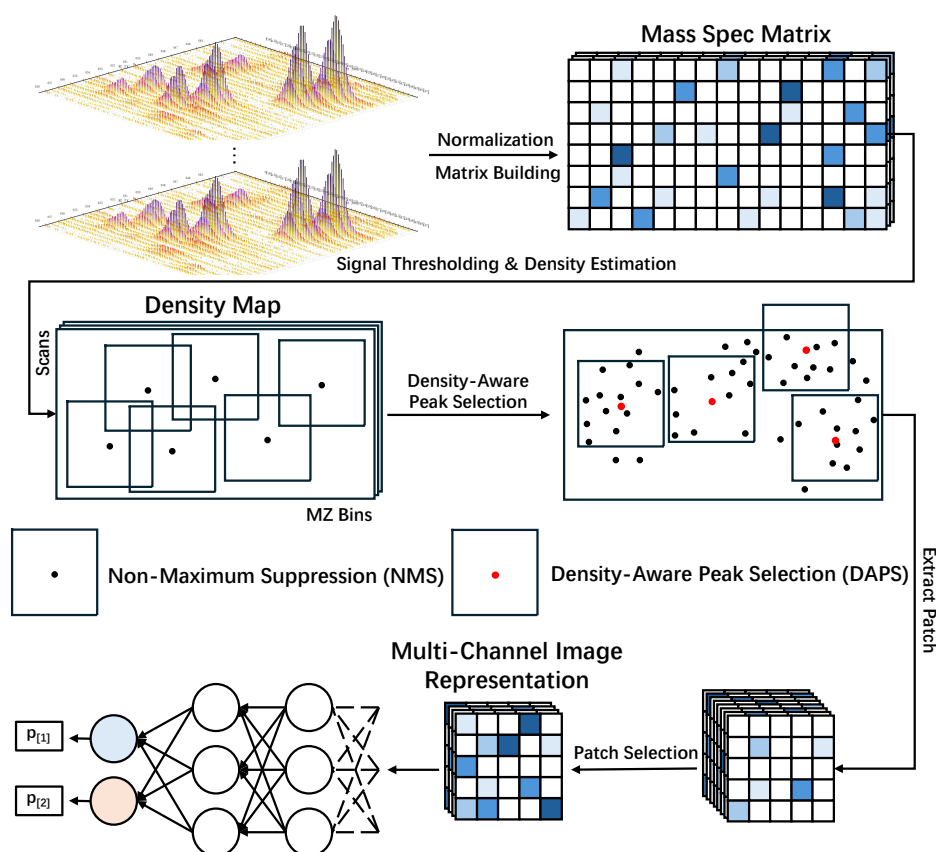
Fundamentally, MS data can be viewed as a two-dimensional data matrix composed of retention time (RT), mass-to-charge ratio ( $m/z$ ), and ion intensity, allowing it to be indirectly treated as an “image” [18]. Utilizing the entire mass spectrum as an “image” for data analysis is an alternative and promising strategy. However, a key engineering challenge is that the dimensions of raw MS matrices are typically enormous. Directly inputting them into deep learning models designed for standard image sizes would lead to an explosion of network parameters. This surge in complexity not only causes catastrophic computational resource consumption but also leads to severe model overfitting. In this context, overfitting means the model memorizes the training data’s noise and specific details instead of learning the core, generalizable patterns. This prevents the model from generalizing to new data, thus limiting its practical utility. Consequently, effective dimensionality reduction or feature selection is imperative before applying deep learning models. To address this challenge, researchers have undertaken a series of pioneering explorations. Shen et al. [18] proposed the deepPseudoMSI method, which converts LC-MS metabolomics data into a fixed-size (e.g.,  $224 \times 224$ ) “pseudo-mass spectrometry image” through a data binning strategy. Specifically, the retention time and  $m/z$  dimensions are divided into a grid, and the intensities of all data points falling within the same grid cell are aggregated to represent the “pixel” intensity. Similarly, Zhang et al. [19] introduced the DIAT data format in proteomics, compressing data by pooling the  $m/z$  dimension and successfully applying it to deep learning models for phenotype classification. While these methods resolve model compatibility issues, their data compression strategies lead to the loss of original mass spectral features. To perform effective feature selection while preserving data characteristics, Wang et al. [20] developed the MetImage method. This approach divides the complete MS matrix into multiple smaller patches using a grid, then selects the most information-rich patches based on entropy or mean values to construct a multi-channel image for input into a deep learning model. This encoding strategy preserves the mass spectral features of the original LC-MS data, enabling the straightforward identification of characteristic metabolites within the key images. This, in turn, helps to elucidate the model’s diagnostic principles, thereby enhancing its interpretability.

Building upon the MetImage framework, this paper proposes a framework for constructing multi-channel mass spectrometry images, named MSIMG. The theoretical foundation of this work is that effective texture feature extraction is critical to the success of texture classification methods [21]. We argue that MetImage has a fundamental flaw in this regard: its fixed-grid partitioning strategy can disrupt the inherent and meaningful texture structures within the data [22], such as splitting a complete signal peak across two or more adjacent patches, thereby compromising feature integrity [23]. To overcome this limitation,

MSIMG introduces a novel “density-peak-centric” patch extraction strategy. This approach, inspired by object detection in computer vision, uses Density Map Estimation [24] and Non-Maximum Suppression (NMS) [25] techniques to precisely locate the centers of signal peaks. These centers then serve as anchors for dynamic, content-aware patch extraction, replacing the fixed-grid partitioning of MetImage that can compromise feature integrity. We hypothesize that this “density-peak-centric” patch extraction strategy can more robustly capture the core features associated with biological phenotypes. This study aims to further explore the feasibility, effectiveness, and potential boundaries of applying computer vision techniques to mass spectrometry data analysis through the MSIMG method.

## 2. Materials and Methods

This study proposes a framework for constructing multi-channel image representations of mass spectrometry data, named MSIMG, designed to convert raw LC-MS data into a multi-channel image format suitable for downstream deep learning classification tasks. The core of MSIMG is a density-based greedy peak selection algorithm, termed Density-Aware Peak Selection (DAPS), which dynamically identifies and extracts the most information-rich data regions. Figure 1 illustrates the complete workflow of MSIMG, from raw data processing to the final construction of the multi-channel image representation.



**Figure 1.** Schematic workflow of the MSIMG method for converting mass spectrometry data into a multi-channel image representation. The diagram illustrates how MSIMG transforms MS data into a multi-channel image. The process mainly includes: preprocessing raw data into a mass spec matrix; applying the DAPS algorithm (combining density map estimation and NMS) to determine candidate patch centers (represented by red dots), which are selected from a larger pool of initial peaks (black dots) identified across the dataset; and finally, selecting the most informative patches based on entropy scores and stacking them to form the final multi-channel image representation. Notably, the DAPS and final patch selection steps are performed on the entire training set to ensure the global representativeness of the features.

## 2.1. Mass Spec Matrix Construction

All raw LC-MS data are first converted into a universal mass spectrometry data format (e.g., mzML) and then transformed into a two-dimensional mass spec matrix through a unified preprocessing pipeline, which mainly consists of  $m/z$  binning and intensity normalization. Each LC-MS file contains a series of  $m/z$  values and their corresponding ion intensities at each retention time point. The purpose of  $m/z$  binning is to align the  $m/z$  axes across different samples and scan times. Specifically, given an  $m/z$  range  $[mz_{\min}, mz_{\max}]$  and a *bin size*, the index  $i$  for any  $m/z$  value is calculated as follows:

$$i = \lfloor \frac{mz - mz_{\min}}{bin\_size} \rfloor \quad (1)$$

All ion intensities falling into the same *bin*  $B_i$  are summed to obtain the final intensity value  $I_{B_i}$  for that bin:

$$I_{B_i} = \sum_{j \text{ s.t. } \lfloor \frac{mz_j - mz_{\min}}{bin\_size} \rfloor = i} Intensity(mz_j) \quad (2)$$

After applying this operation to the mass spectra at all retention time points, each LC-MS file is converted into a two-dimensional mass spec matrix  $M$  with dimensions  $(N_{scans}, N_{bins})$ , where  $N_{scans}$  is the number of retention time points and  $N_{bins}$  is the total number of bins along the  $m/z$  axis.

To eliminate variations in overall signal intensity between samples and adapt the data to the input range of deep learning models, the constructed mass spec matrix  $M$  is normalized. Specifically, the 99.9th percentile of all non-zero elements in matrix  $M$  is set as the maximum value, and the minimum is set to 0. The normalized values in the matrix are then clipped to the range  $[0,1]$  to ensure numerical stability, yielding the final normalized mass spec matrix  $M_{norm}$ .

## 2.2. Density-Aware Peak Selection Strategy

The core of the MSIMG method lies in its data-driven strategy for generating candidate patch locations, the DAPS algorithm. This algorithm is designed to identify the most representative shared signal hotspot regions across the entire dataset and generate a set of candidate patch center coordinates. It discards the fixed-grid partitioning strategy in favor of dynamic, “information-density-centric” patch extraction.

The MSIMG method first processes each sample in the training set to generate a density map and extract a set of candidate peak points. The purpose of generating a density map is to identify potentially biologically significant signal-dense regions from the sparse and noisy mass spec matrix. This process aims to convert discrete signal points into a continuous density surface, thereby highlighting signal “hotspots” in the matrix that guide subsequent patch location selection. Specifically, an intensity threshold  $\tau_{intensity}$  is set. Pixels in the normalized mass spec matrix  $M_{norm}$  with intensity values greater than this threshold are considered effective signal points, while the rest are treated as background. This operation converts the continuous intensity matrix  $M_{norm}$  into a binary signal mask matrix  $S$ . After obtaining the signal mask matrix  $S$ , a sliding window of size  $w \times w$  is convolved over  $S$  to estimate the local signal density at each pixel, thus generating a density map  $D$ . Each pixel  $(i,j)$  in the density map  $D$  is the sum of all signal mask values within the  $w \times w$  window centered at that point. The process can be described by the following equation:

$$D_{i,j} = \sum_{x,y \in \text{Window}_{i,j}} S_{x,y} \quad (3)$$

The density map  $D$  is a 2D array of the same size as the mass spec matrix, where higher values indicate denser effective signals. However, bright regions (i.e., signal-dense areas) on the density map  $D$  may contain multiple adjacent local maxima. To precisely locate a single, most representative peak from each signal-dense region, MSIMG applies the Non-Maximum Suppression (NMS) algorithm. Originating from the field of object detection in computer vision, NMS is a key post-processing technique for eliminating redundant detection boxes [26]. Its core mechanism of selecting a single optimal target is well-suited for our peak localization task. In the MSIMG method, we adapt its core idea and implement peak selection through an iterative greedy strategy. Specifically, at the beginning of each iteration, the point with the global maximum density,  $p_{max}$ , is located in the current density map. If its density value is above a predefined density threshold  $\tau_{density}$ , it is recorded as a valid candidate peak. Simultaneously, to prevent re-selecting peaks from the same dense region in subsequent iterations, all density values within a rectangular area centered at  $p_{max}$  (with the same size as the predefined patch size) are set to 0. This process continues until a preset number of candidate peaks are extracted or no points satisfying the density threshold can be found.

Through the above process, MSIMG extracts a set of high-quality, spatially separated candidate peak points from a single sample, which accurately represent the centers of the most signal-dense regions in that sample. This procedure is then repeated for every sample in the training set, and all extracted candidate peak points are aggregated to form a global set of candidate peaks,  $P_{all}$ . To select a final set of patch centers from this dense point set that best represents the dataset's characteristics without spatial overlap, we designed a greedy selection algorithm—DAPS. This algorithm iteratively selects an “optimal peak” that maximizes local candidate peak coverage. To perform spatial queries efficiently, the global candidate peak set  $P_{all}$  is first structured into a k-d tree [27]. In each iteration, DAPS evaluates the “potential value” of a subset of currently available candidate peaks. This value is defined as the number of candidate peaks contained within a circular region centered at a candidate peak  $p_i$ . This region is defined as a circle  $B_{p_i,r}$  with a radius  $r$  equal to the distance from the patch center to its diagonal corner. Therefore, the potential value,  $\text{Score}_{p_i}$ , of a candidate peak  $p_i$  can be formalized as:

$$\text{Score}_{p_i} = |\{p_j | p_j \in P_{avail} \cap B_{p_i,r}\}| \quad (4)$$

where  $P_{avail}$  is the set of all currently available peaks and  $|\cdot|$  denotes the number of elements in the set. DAPS then selects the peak with the highest score,  $p_{best}$ , as the optimal choice for the current iteration:

$$p_{best} = \underset{p_i \in P_{cand}}{\operatorname{argmax}} \text{Score}_{p_i} \quad (5)$$

Once  $p_{best}$  is selected, it is added to the final set of patch center coordinates,  $P_{selected}$ . Concurrently, all candidate peaks falling within its coverage area  $B_{p_{best},r}$  are marked as unavailable to ensure that subsequent selections do not overlap. This iterative process continues until the density of the remaining available peaks falls below the preset threshold  $\tau_{min\_peaks}$ . DAPS ultimately outputs a set of candidate patch center coordinates,  $P_{selected}$ , that maximizes signal coverage across the entire training set.

### 2.3. Construction of the Multi-Channel Image Representation

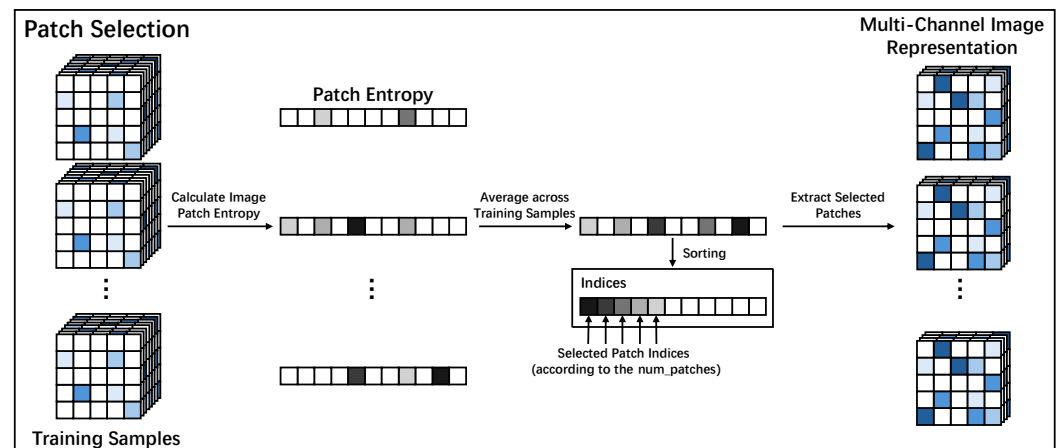
After generating the final set of patch center coordinates,  $P_{selected}$ , using the DAPS algorithm, the next step is to construct the final multi-channel image representation for each sample. This is achieved through a unified, entropy-based scoring and ranking process based on these candidate locations. The process begins by extracting candidate patches for scoring. For each sample in the training set, MSIMG extracts patches from its corresponding

normalized matrix  $M_{norm}$  according to the candidate center coordinates in  $P_{selected}$ . Specifically, for each center coordinate  $(r, c)$  in  $P_{selected}$ , a patch of size  $patch\_height \times patch\_width$  centered at that point is extracted. If the extraction area exceeds the matrix boundaries, it is padded with a preset value (e.g., 0). After this step, for each candidate location, we have the corresponding patch instances from all samples.

Next, MSIMG employs one-dimensional information entropy as the evaluation metric for the final patch selection process. For each candidate patch location, every sample in the training set has a corresponding patch. However, the amount of information contained in patches at the same location varies across different samples, resulting in different entropy values. Therefore, to select the locations that are most representative and information-rich across the entire training set, we first calculate the average information entropy for each candidate location across all training samples. The patch locations with the highest average information entropy are then chosen for the final selection. The entire workflow for this patch selection process is illustrated in detail in Figure 2. The core of this process is the calculation of information entropy. Specifically, for a given patch  $P$ , its information entropy  $H(P)$  is defined as:

$$H(P) = - \sum_{k=0}^{255} p_k \log_2(p_k) \quad (6)$$

where  $p_k$  is the proportion of pixels with intensity value  $k$  (normalized to the 0–255 range) in patch  $P$ .

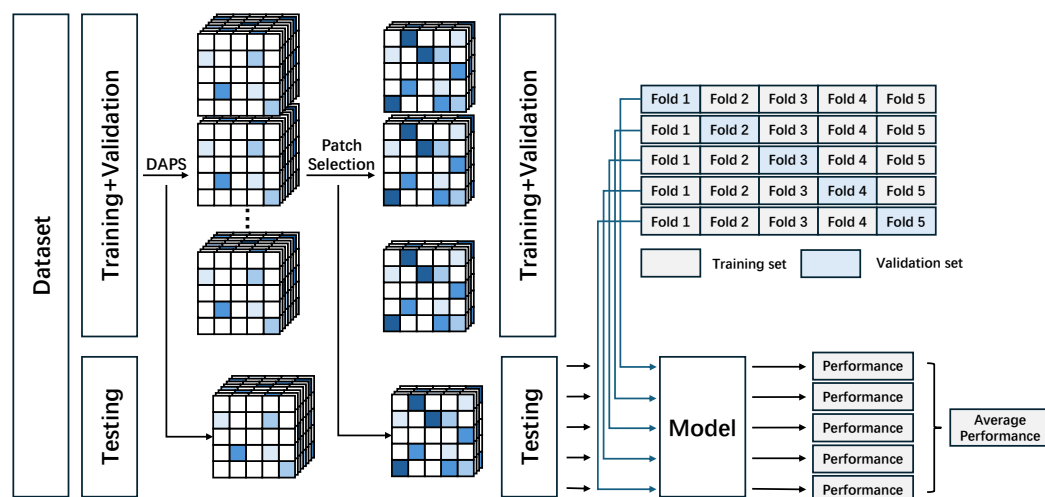


**Figure 2.** The Entropy-based patch selection workflow in MSIMG. For each candidate patch location, information entropy is calculated for the corresponding patch in every training sample. These values are then averaged across the training set to produce a single representative score for each location. The locations are ranked by their average entropy, and the top  $num\_patches$  are selected. Patches from these top-ranked locations are extracted from every sample in the dataset and stacked to form the final multi-channel image representation.

After calculating the average information entropy for all candidate patch locations, the top-ranking  $num\_patches$  locations are selected as the “golden locations” for constructing the multi-channel image representation for all samples. Specifically, for every sample in the dataset (including both training and testing sets), MSIMG selects the patches at these  $num\_patches$  golden locations and stacks them along the channel dimension. This forms a three-dimensional tensor with dimensions  $(num\_patches, patch\_height, patch\_width)$ . This 3D tensor is the “multi-channel image” representation of the sample and can be directly used as input for a CNN model, where  $num\_patches$  corresponds to the number of input channels for the model.

## 2.4. Model Training and Evaluation Strategy

To ensure a robust and unbiased evaluation of the model's performance, we designed a rigorous training and evaluation pipeline. As shown in Figure 3, the entire dataset is first divided into a training+validation set and an independent hold-out test set using an 8:2 stratified sampling split. On the training+validation set, we employed a 5-fold cross-validation strategy to train and select the model. For each fold, the model is trained on the training data and its performance is monitored on the validation data. All final performance evaluations are conducted on the hold-out test set to ensure the generalizability of the results.



**Figure 3.** Model training and evaluation workflow of the MSIMG method. The dataset is first split into a training+validation set and an independent hold-out test set. The proposed DAPS algorithm and the subsequent patch selection process are executed only on the training+validation set to determine a set of the most informative patch locations. This selected set of locations is then applied to the test set to construct its corresponding multi-channel image representations, thus ensuring the independence of the test data.

To enhance model generalization and prevent overfitting, we applied data augmentation strategies exclusively to the training data. These included a series of random image transformations designed to simulate variations encountered during actual MS data acquisition, such as random affine transformations, random intensity scaling, addition of Gaussian noise, and random erasing of image regions. These augmentations ensure that the model learns more robust features that are insensitive to minor perturbations.

For model optimization, the Adam optimizer was used for efficient parameter updates. To prevent overfitting, L2 regularization (weight decay) was applied to penalize model complexity, helping the model learn more generalizable feature representations. To dynamically adjust the learning rate, we used a ReduceLROnPlateau scheduler: the learning rate was multiplied by 0.1 when the validation loss did not improve for 5 consecutive epochs. Furthermore, to address class imbalance in the dataset, we calculated class weights and incorporated them into the cross-entropy loss function. The entire training process was run for 64 epochs with a batch size of 8. We also implemented an early stopping mechanism: if the validation loss failed to improve for 10 consecutive epochs, training was terminated early, and the model weights with the best performance were saved. After completing the 5-fold cross-validation, we obtained five independently trained models. Each of these five models was then evaluated on the independent hold-out test set, and the mean and confidence interval of various performance metrics (such as accuracy, F1-score, etc.) were calculated and reported as the final performance of our method. All experiments

were conducted using Python 3.9.23 and PyTorch 2.7.1 on a workstation equipped with an NVIDIA GeForce RTX 4070 Ti Super GPU.

### 2.5. Datasets

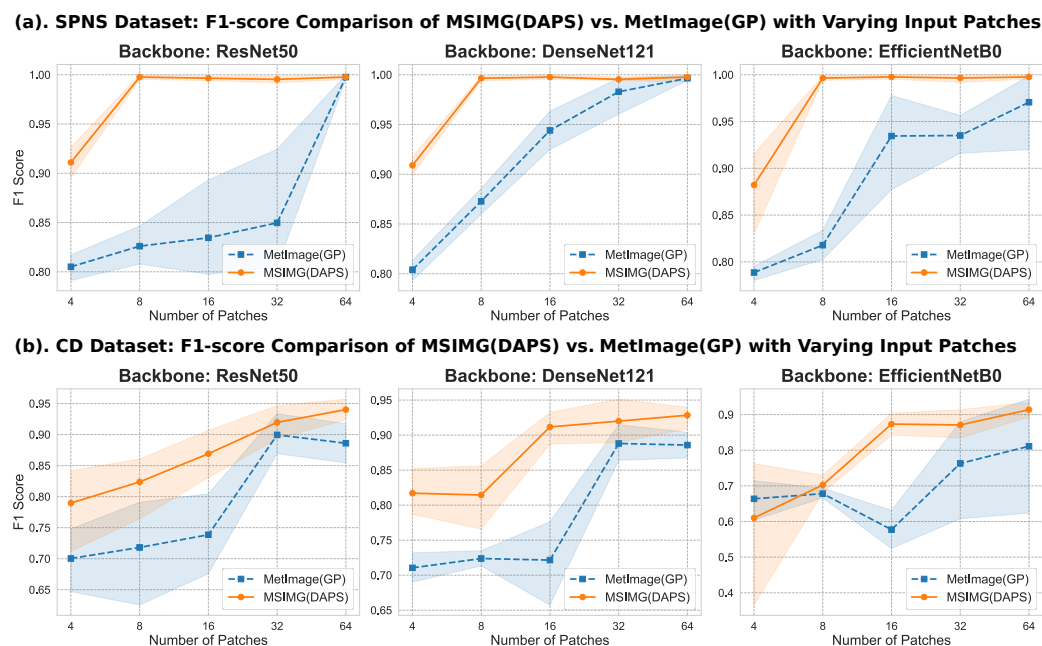
We conducted experiments on two independent, publicly available mass spectrometry datasets (SPNS and CD) to evaluate the effectiveness of our proposed MSIMG method. The SPNS dataset originates from a metabolomics and lipidomics study on Solitary Pulmonary Nodules (SPNs) [28]. The study found significant pattern differences in circulating metabolites and lipids between SPN patients (both benign and malignant) and healthy controls, but no significant differences between benign and malignant SPNs. Given this, and considering that distinguishing between benign and malignant nodules is beyond the scope of this study, we grouped the benign and malignant SPN samples into a single class. This created a binary classification task to distinguish SPN patients from healthy controls, comprising 880 SPN samples and 279 healthy control samples. The CD dataset is from a metabolomics study investigating the relationship between the host and gut microbiota in Crohn's Disease (CD). This dataset includes a total of 294 fecal samples, with 250 from CD patients and 44 from healthy controls. These two datasets, representing different disease types and sample origins, provide a solid foundation for evaluating the performance and generalization capability of our method.

## 3. Results

This study compares the MSIMG method against two baseline methods from different paradigms. The first is the traditional peak list method, which represents the standard workflow in mass spectrometry analysis. This involves extracting feature peaks using common MS processing software, then converting the peak list into a one-dimensional feature vector for input into a classification model. The second baseline is the grid-based MetImage method, which uses a fixed-size sliding window to scan across the entire mass spec matrix at a fixed stride. This generates a corresponding set of candidate patches at fixed grid locations for each sample in the dataset. In our study, both the MSIMG and MetImage methods use the same subsequent process for patch selection and multi-channel image construction to ensure controlled experimental variables.

### 3.1. Parameter Sensitivity Analysis: Selection of Multi-Channel Image Dimensions and Number of Channels

In methods like MSIMG that represent mass spectrometry data as a multi-channel image, the number of selected patches, `num_patches`, directly determines the channel count of the final image and is a critical hyperparameter affecting model performance. To thoroughly investigate the impact of the number of input patches on model performance and to compare the feature selection efficiency of our proposed MSIMG(DAPS) against the baseline MetImage(GP), we conducted a series of parameter sensitivity experiments. Specifically, on both the SPNS and CD datasets, we systematically evaluated how the classification performance of downstream models was affected by an increasing number of input patches, using three different CNN backbone models (ResNet [29], DenseNet [30], and EfficientNet [31]). It is worth noting that the maximum number of candidate patches MSIMG can extract is determined by the signal density and distribution characteristics of the dataset itself; we considered these factors when setting the parameter range for our experiments.

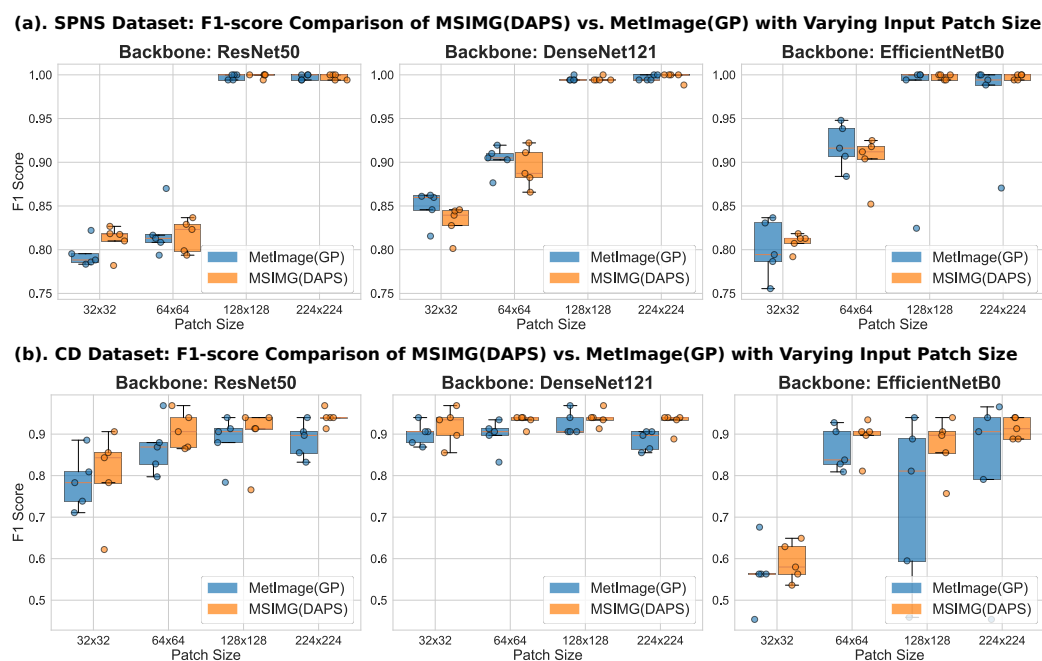


**Figure 4.** F1-score comparison between MSIMG(DAPS) and MetImage(GP) with varying numbers of input patches. This figure shows the impact of the number of selected patches on model performance with a fixed patch size of 224. (a) The performance trends on the SPNS dataset. (b) The performance trends on the CD dataset. For both datasets, the comparison was performed across three different CNN backbones (ResNet50, DenseNet121, and EfficientNetB0). The lines represent the mean F1-score across 5-fold cross-validation, and the shaded areas indicate the 95% confidence intervals.

The results of the patch number sensitivity analysis, as shown in Figure 4, reveal a clear and consistent trend: for both methods, increasing the number of patches generally leads to improved model performance across all tested conditions (with a fixed patch size of 224). This suggests that more patches provide richer information, helping the model learn more discriminative features. However, the two methods show significant differences in their performance improvement trajectories. Our proposed MSIMG method outperforms the MetImage method under almost all parameter settings and reaches its performance saturation point much earlier. For instance, on the SPNS dataset, regardless of the backbone model used, the DAPS strategy of MSIMG achieves optimal performance with an F1-score approaching or exceeding 0.95 using only 8 carefully selected patches. In contrast, the MetImage method still shows a considerable performance gap at 16 patches and requires many more patches to improve slowly. Furthermore, we observed that after MSIMG reaches its optimal performance, the F1-score curve flattens, with subsequent minor fluctuations being statistically insignificant and more likely due to experimental variance than a performance decline. This indicates that the DAPS strategy of MSIMG can efficiently identify the most representative regions in the dataset, thereby extracting key discriminative information.

In addition to the number of input patches, the patch size is another critical hyperparameter, as it determines the local information range the model can perceive in a single observation. To investigate the impact of this parameter, we systematically evaluated the effect of four different sizes on the classification performance of MSIMG(DAPS) and MetImage(GP), while keeping the number of patches fixed at 64. The results, presented in Figure 5, clearly demonstrate a non-linear relationship between patch size and model performance. A general observation is that excessively small patch sizes lead to a significant drop in performance. As the patch size increases, model performance tends to stabilize.

This trend is particularly evident with the EfficientNetB0 model on the CD dataset, where performance for both methods jumps substantially when the patch size increases from 32 to 64, followed by a plateau as the size further increases from 64 to 224. This suggests that patches that are too small may fail to capture a complete, discriminative signal peak pattern, thus limiting the model's learning capacity. Once the patch size is sufficient to contain key feature information, further increases do not yield significant performance gains. Considering performance stability and compatibility with mainstream CNN model inputs, we determined that 224 is a suitable size and used it in all subsequent comparative experiments.



**Figure 5.** F1-score comparison between MSIMG(DAPS) and MetImage(GP) with varying input patch sizes. This figure shows the impact of patch size on model performance with a fixed number of 64 input patches. The performance comparison is shown for (a) the SPNS dataset and (b) the CD dataset. For both datasets, the experiments were conducted using three different convolutional neural network backbones (ResNet50, DenseNet121, and EfficientNetB0) across four different patch sizes. The box plots represent the quartile range of results from 5-fold cross-validation, with the center line indicating the median.

This series of experiments leads to an important conclusion: both MetImage and MSIMG can be viewed as feature selection strategies, aiming to select the most discriminative subset of features (i.e., patches) from the high-dimensional mass spec matrix. The final classification performance is theoretically capped by the intrinsic information content of the dataset and the model architecture. The core difference between the two methods lies in the efficiency with which they approach this performance limit. The GP strategy, through a near-exhaustive approach, can eventually approach this limit by including a sufficient number of informative regions. In contrast, our proposed MSIMG method, with its data-driven, “density-peak-centric” intelligent selection strategy, can accurately locate and select these critical discriminative regions from the outset. This allows the DAPS strategy to reach the performance ceiling faster and more efficiently, with fewer patches than the GP strategy. Overall, the performance of the DAPS strategy not only demonstrates its effectiveness but also strongly validates that applying advanced, content-aware computer vision strategies to mass spectrometry data analysis is a promising and effective approach.

### 3.2. Performance Comparison of Multi-Channel Image Representation and Traditional Methods

To comprehensively evaluate the effectiveness of our proposed MSIMG multi-channel image representation method, we conducted a direct performance comparison with the traditional peak list method. Based on the preceding parameter sensitivity analysis, the MSIMG method was configured with a patch size of  $224 \times 224$  and a patch number of 64. For experimental design, we equipped the traditional peak list method with a variety of classifiers. For the classic machine learning models, we utilized the implementations from the Python scikit-learn library to establish a performance baseline. Specifically, the Random Forest was used with its default hyperparameters, including 100 estimators. The Support Vector Machine was configured with a Radial Basis Function (RBF) kernel, while other parameters like regularization were kept at their default values. The Linear Discriminant Analysis was also used with its default settings. For all of these classical models, no further hyperparameter tuning was performed. In addition to these models, we also used three convolutional neural networks (ResNet50, DenseNet121, and EfficientNetB0) adapted for one-dimensional data. Correspondingly, for the two-dimensional multi-channel images generated using the MSIMG method, we used the standard 2D versions of the same three CNN models to ensure a fair and consistent comparison.

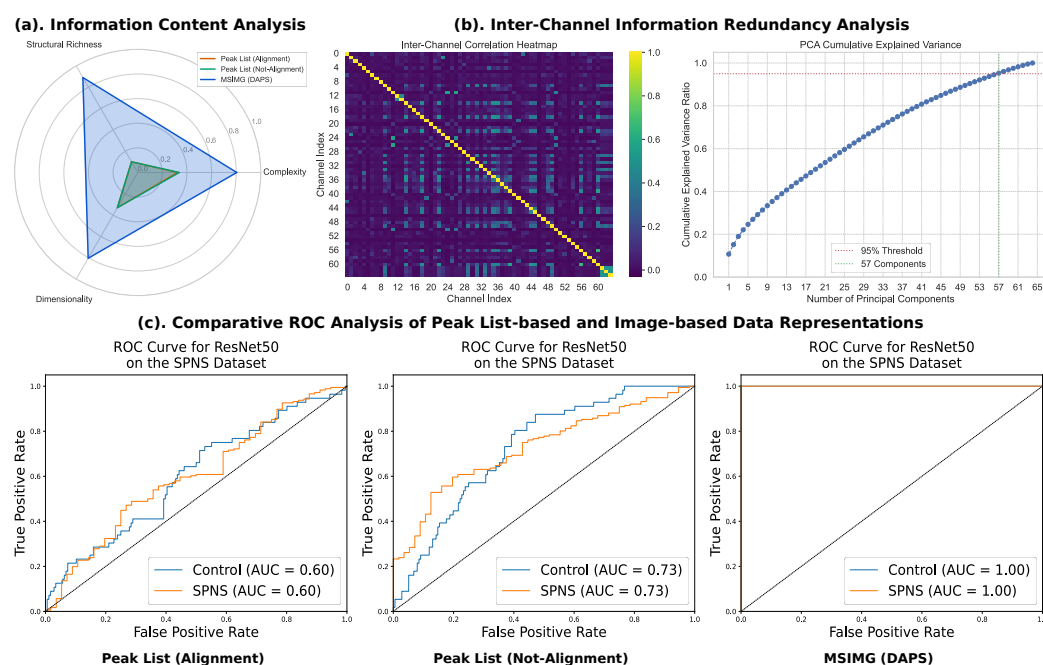
We conducted extensive comparative experiments on the SPNS and CD datasets, with the results presented in Tables 1 and 2. The results clearly show that the MSIMG multi-channel image representation method comprehensively outperforms the traditional peak list method in terms of performance. On the larger SPNS dataset, this difference in performance is particularly stark. With the peak list method, both classic machine learning models and 1D deep learning models exhibited limited performance, with the highest F1-score reaching only 0.6671. In sharp contrast, the MSIMG method approached near-perfect performance across all three CNN models, with F1-scores consistently around 0.9977. This indicates that MSIMG provides a feature representation with extremely high information density and discriminative power, something that the one-dimensional feature vectors generated by traditional peak list methods struggle to achieve. A similar trend was observed on the smaller CD dataset. Although the peak list method performed reasonably well on this dataset, achieving a maximum F1-score of 0.7695, the MSIMG method still demonstrated a definitive advantage, with all three models surpassing an F1-score of 0.91, and the ResNet50 model reaching a high of 0.9402.

**Table 1.** Performance comparison of MSIMG and peak list methods on the SPNS dataset.

| Model          | Peak List (Align) |            | Peak List (Not-Align) |            | MSIMG      |            |
|----------------|-------------------|------------|-----------------------|------------|------------|------------|
|                | Accuracy          | F1 Score   | Accuracy              | F1 Score   | Accuracy   | F1 Score   |
| RF             | 0.7819            | 0.5563     | 0.8129                | 0.6671     | -          | -          |
|                | $\pm 0.01$        | $\pm 0.02$ | $\pm 0.01$            | $\pm 0.02$ | -          | -          |
| SVM            | 0.7604            | 0.4387     | 0.7595                | 0.4351     | -          | -          |
|                | $\pm 0.00$        | $\pm 0.02$ | $\pm 0.00$            | $\pm 0.01$ | -          | -          |
| LDA            | 0.5414            | 0.4905     | 0.5586                | 0.5126     | -          | -          |
|                | $\pm 0.02$        | $\pm 0.02$ | $\pm 0.03$            | $\pm 0.03$ | -          | -          |
| ResNet50       | 0.7181            | 0.4745     | 0.6155                | 0.5694     | 0.9983     | 0.9977     |
|                | $\pm 0.05$        | $\pm 0.05$ | $\pm 0.03$            | $\pm 0.02$ | $\pm 0.00$ | $\pm 0.00$ |
| DenseNet121    | 0.6991            | 0.4699     | 0.6500                | 0.5966     | 0.9983     | 0.9977     |
|                | $\pm 0.08$        | $\pm 0.05$ | $\pm 0.02$            | $\pm 0.02$ | $\pm 0.00$ | $\pm 0.01$ |
| EfficientNetB0 | 0.7104            | 0.4806     | 0.6595                | 0.5988     | 0.9983     | 0.9977     |
|                | $\pm 0.06$        | $\pm 0.05$ | $\pm 0.04$            | $\pm 0.03$ | $\pm 0.00$ | $\pm 0.00$ |

**Table 2.** Performance comparison of MSIMG and peak list methods on the CD dataset.

| Model          | Peak List (Align) |            | Peak List (Not-Align) |            | MSIMG      |            |
|----------------|-------------------|------------|-----------------------|------------|------------|------------|
|                | Accuracy          | F1 Score   | Accuracy              | F1 Score   | Accuracy   | F1 Score   |
| RF             | 0.8542            | 0.5689     | 0.9085                | 0.7695     | -          | -          |
|                | $\pm 0.02$        | $\pm 0.05$ | $\pm 0.02$            | $\pm 0.05$ | -          | -          |
| SVM            | 0.8440            | 0.5298     | 0.8780                | 0.6424     | -          | -          |
|                | $\pm 0.01$        | $\pm 0.04$ | $\pm 0.01$            | $\pm 0.06$ | -          | -          |
| LDA            | 0.7593            | 0.6760     | 0.8542                | 0.7580     | -          | -          |
|                | $\pm 0.04$        | $\pm 0.04$ | $\pm 0.04$            | $\pm 0.05$ | -          | -          |
| ResNet50       | 0.7424            | 0.6238     | 0.8542                | 0.7677     | 0.9661     | 0.9402     |
|                | $\pm 0.06$        | $\pm 0.04$ | $\pm 0.03$            | $\pm 0.04$ | $\pm 0.01$ | $\pm 0.02$ |
| DenseNet121    | 0.7864            | 0.6817     | 0.8237                | 0.7556     | 0.9593     | 0.9284     |
|                | $\pm 0.16$        | $\pm 0.16$ | $\pm 0.09$            | $\pm 0.10$ | $\pm 0.02$ | $\pm 0.02$ |
| EfficientNetB0 | 0.8339            | 0.7035     | 0.8237                | 0.7355     | 0.9492     | 0.9138     |
|                | $\pm 0.03$        | $\pm 0.04$ | $\pm 0.09$            | $\pm 0.08$ | $\pm 0.02$ | $\pm 0.03$ |



**Figure 6.** Comprehensive analysis linking data representation to classification performance. This figure evaluates three data preprocessing methods: Peak List (Alignment), Peak List (Not-Alignment), and our proposed MSIMG (DAPS). **(a)** Information Content Analysis via a radar chart comparing the three methods across normalized metrics of Complexity, Structural Richness, and Dimensionality. The MSIMG (DAPS) method exhibits substantially higher structural richness and dimensionality. **(b)** Inter-channel information redundancy analysis for the 64-channel MSIMG (DAPS) representation. The correlation heatmap (left) indicates minimal linear correlation between channels. The PCA cumulative explained variance curve (right), where the blue line plots the cumulative explained variance against the number of principal components, confirms low data redundancy, requiring 57 components to explain 95% of the total variance. **(c)** Comparative ROC curves for a ResNet50 classifier on the SPNS vs. Control classification task. The model trained on the MSIMG (DAPS) representation achieves a perfect Area Under the Curve (AUC) of 1.00, significantly outperforming models trained on the Peak List (Not-Alignment) (AUC = 0.73) and Peak List (Alignment) (AUC = 0.60) representations.

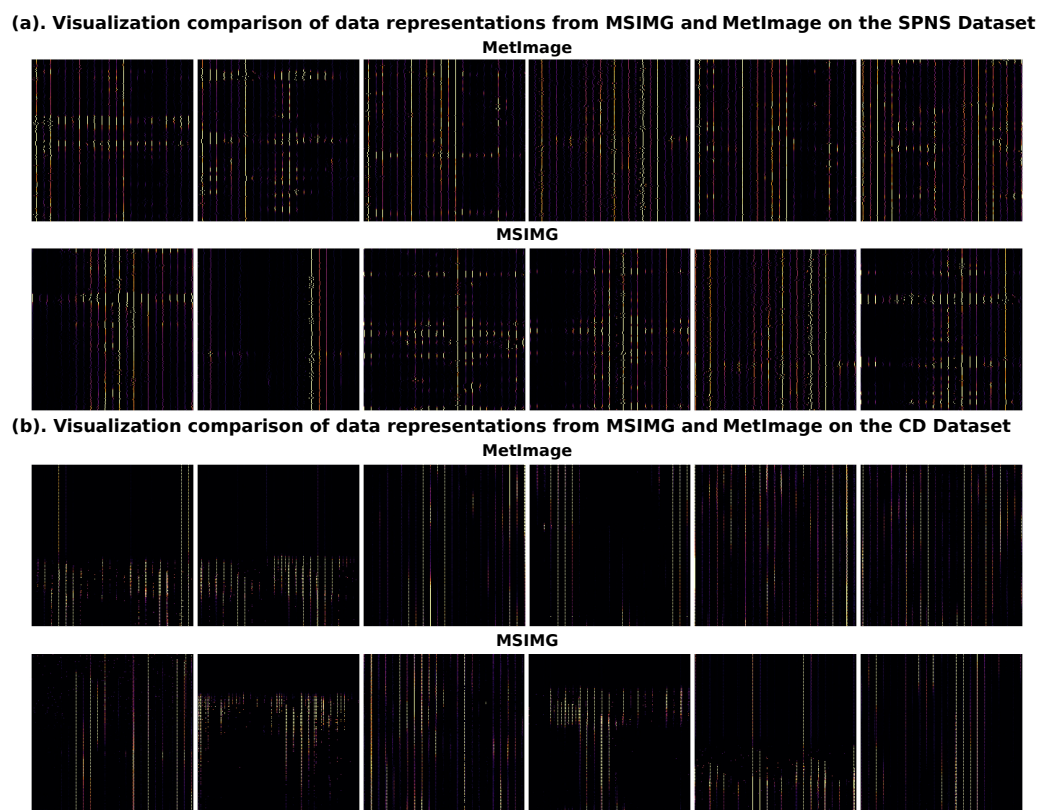
To further investigate the underlying reasons for the superior performance of our method, we conducted a comprehensive analysis of the intrinsic properties of the different data representations, with the results visualized in the Figure 6. The information content

analysis reveals that the MSIMG (DAPS) representation possesses substantially greater Structural Richness and Dimensionality compared to both peak list methods, indicating that our approach captures a more complex and comprehensive feature space. Furthermore, we assessed the information redundancy within the 64-channel MSIMG representation. The low inter-channel correlation shown in the heatmap, combined with the fact that 57 principal components are required to explain 95% of the total variance, provides strong evidence that each channel contributes unique, non-redundant information. This highlights the efficiency of our DAPS selection strategy in creating an information-dense representation. Crucially, this superior data quality translates directly into classification power, as demonstrated by the comparative ROC analysis. The ResNet50 model trained on MSIMG data achieves a perfect Area Under the Curve (AUC) of 1.00, dramatically outperforming the models trained on the not-aligned (AUC = 0.73) and aligned (AUC = 0.60) peak lists. Collectively, this analysis confirms that the MSIMG framework provides a superior input paradigm for deep learning classifiers. Its success stems from a data-driven patch selection strategy that effectively captures and preserves the intrinsic structural and textural information within the data. By “encoding” this information into an image-based representation optimized for visual models, powerful CNNs can fully leverage their pattern recognition strengths, leading to breakthrough performance in mass spectrometry data analysis.

### 3.3. Visualization of the Effectiveness of Multi-Channel Image Representation

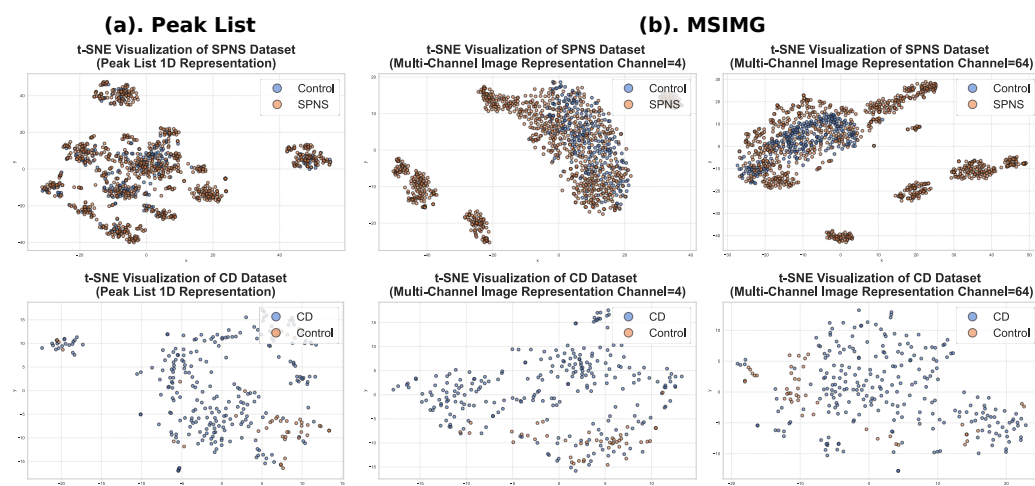
While the performance metrics in the preceding sections establish the superiority of the MSIMG method, we conducted further visualization analyses to intuitively understand the underlying reasons for this superior performance. This section aims to provide a direct visual comparison and evaluation from two key perspectives: the quality of the patches used as model inputs and the separability of the learned features that serve as model outputs. This will clarify how the data-driven strategy of the MSIMG method translates into an enhanced final classification performance.

To investigate the source of MSIMG’s feature advantage, we directly compared the visual quality of patches selected by our method with those selected by the baseline MetImage method. Figure 7 displays example patches (from channels 11 to 16) extracted by both methods on the SPNS and CD datasets, with a patch size of  $224 \times 224$ . There is a significant difference in the content of the patches generated by the two methods. Compared to MetImage’s fixed-grid partitioning strategy, the patches selected by MSIMG generally contain clear and continuous signal peak texture structures; particularly on the SPNS dataset, the extracted patches exhibit distinct clustered textures. In contrast, the content of patches selected by the MetImage strategy appears more random, often capturing either background noise with no discernible pattern or incomplete signal fragments that have been arbitrarily cut by the grid. For example, on the CD dataset, it is clear that its grid partitioning disrupts the integrity of complete mass spectral peak textures. These stark visual contrasts stem from MSIMG’s “density-peak-centric” strategy, which ensures that the selected regions are signal “hotspots” rather than random background areas. These high-quality, information-dense patches present clear visual patterns that reflect underlying, biologically meaningful molecular signals, providing the model with authentic biological patterns to learn from and forming the solid foundation for MSIMG’s superior performance.



**Figure 7.** Visualization comparison of data representations from MSIMG and MetImage on the SPNS and CD dataset. The figure compares representative patches selected from (a) the SPNS dataset and (b) the CD dataset. For each dataset, the top row shows six feature patches (channels 11–16) generated by the grid-based MetImage method, while the bottom row shows the corresponding patches selected by the density-aware MSIMG method.

An effective way to evaluate the quality of features extracted by different methods is to observe their distribution in a high-dimensional space. In this study, we used the t-SNE dimensionality reduction technique to compare the deep features represented by the 1D vectors from the traditional peak list method with those from the multi-channel images generated by the MSIMG method (after pooling and flattening). Figure 8 shows a t-SNE visualization comparison of the two methods on the SPNS and CD datasets. The subplots to the left of the dashed line clearly show that for the SPNS dataset, the feature vectors from the traditional peak list method result in severe mixing of sample points from different classes in the 2D space, making them nearly indistinguishable and indicating weak feature discriminability. In contrast, the features generated by the MSIMG method demonstrate vastly superior performance. As the number of input patches increases from 4 to 16, and then to 64, the class clusters in the t-SNE plots become progressively clearer and more compact. When 64 patches are used, the samples from different classes form almost completely separate and distinct clusters. This result provides strong evidence that the MSIMG multi-channel image representation can capture far richer and more precise discriminative information than traditional 1D vectors, thereby providing the downstream deep learning model with a high-quality, easily separable feature space.



**Figure 8.** Comparison of the separability of features extracted by MSIMG and the traditional peak list method. This figure uses the t-SNE dimensionality reduction technique to visualize and compare the separability of features extracted by different methods. Subplot (a) shows the sample distribution based on the traditional Peak List 1D representation. Subplot (b) shows the sample distribution for the MSIMG multi-channel image representation, demonstrating improved class separation as the number of channels increases. The top and bottom rows show the results for the SPNS and CD datasets, respectively.

#### 4. Discussion

The experimental results of this study demonstrate that our proposed MSIMG method, by converting raw mass spectrometry data into a multi-channel image representation, achieves significantly superior classification performance compared to the traditional peak list method. This finding not only validates the effectiveness of the MSIMG method but, more importantly, prompts a re-examination of a fundamental issue in mass spectrometry data analysis: the relationship between data representation and the analytical objective.

In traditional mass spectrometry data analysis, a peak list is typically generated for subsequent analytical tasks. The fundamental purpose of this process is to extract and interpret biologically meaningful information from complex, raw instrumental data [32]. A peak list is a structured, denoised, and information-compressed representation designed for human scientists. It discretizes continuous spectral data into a series of identifiable and annotatable metabolites or peptides for pathway analysis, with the ultimate goal of biological interpretation. However, when the analytical objective shifts from biological interpretation to a pure classification or prediction task, it is crucial to question whether this data representation, optimized for biological insight, remains the best choice for achieving optimal performance.

Deep learning models, particularly Convolutional Neural Networks (CNNs), derive their power not from understanding structured tables but from their exceptional ability to perceive spatial patterns, textures, and local correlations in images [33]. From this perspective, a traditional peak list represents a structurally mismatched input format for a CNN, with a significant loss of information. The success of MSIMG demonstrates that breakthrough performance can be achieved by redesigning the data representation to align with the demands of the task (i.e., classification) and the characteristics of the model. By converting mass spectrometry data into an image representation, MSIMG preserves the rich two-dimensional spatial information of the original data, providing the model with a more “native” and information-dense input, thereby achieving the classification goal more effectively.

This brings to mind the classic computer science principle of “Garbage In, Garbage Out.” Typically, this principle means that if the input data itself is of low quality, no model can produce reliable results [34]. However, our research reveals another facet of this principle: even information-rich raw data can become “garbage input” for a downstream model if it is inappropriately represented [35]. On the SPNS and CD datasets, the poor performance of the traditional peak list method was not due to low-quality raw mass spectrometry data, but because the one-dimensional vector representation produced after data filtering had lost a substantial amount of information, becoming an “information-poor” input. MSIMG provides a superior data representation strategy. It transforms high-quality raw mass spectrometry data into a high-quality model input through an appropriate representation method, thereby achieving satisfactory results. Notably, this data representation paradigm, optimized for Convolutional Neural Networks, may in turn offer new possibilities for enhancing the efficiency of extracting biological insights. Traditional peak list generation relies on fixed algorithms and parameters, which may overlook important low-abundance signals. In contrast, MSIMG employs an unbiased, data-driven process to precisely locate the most discriminative information-rich regions for the classification task. These patches, “attended to” by the model, are essentially signal hotspots that can guide researchers to focus on specific mass-to-charge ratio-retention time ( $m/z$ -RT) windows for more in-depth targeted analysis and identification. This strategy is expected to significantly enhance the efficiency of discovering potential biomarkers, thereby establishing an effective human-computer collaborative analysis loop.

## 5. Conclusions

This study introduces a novel framework named MSIMG, which transforms raw mass spectrometry data into multi-channel images for deep learning-based phenotype classification. By employing a data-driven, “peak-centric” patch selection strategy, MSIMG effectively preserves the integrity of key signal patterns, overcoming the critical limitations of traditional grid-based methods. Our extensive experiments demonstrate that the MSIMG method significantly improves classification performance across multiple datasets and deep learning architectures compared to traditional peak list methods. These results highlight the profound impact of data representation on model performance and validate the effectiveness of applying content-aware computer vision strategies to analytical chemistry data. This work paves the way for future research into more robust and interpretable deep learning models for clinical proteomics and metabolomics.

**Author Contributions:** Conceptualization, F.Z.; methodology, X.X. and F.Z.; software, F.Z.; validation, X.X. and F.Z.; formal analysis, X.X. and F.Z.; investigation, F.Z. and L.G.; resources, X.X.; data curation, Y.W., L.G. and W.Z.; writing—original draft preparation, F.Z.; writing—review and editing, X.X. and F.Z.; visualization, F.Z.; supervision, X.X. and B.G.; project administration, X.X.; funding acquisition, X.X. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Science & Technology Fundamental Resources Investigation Program (Grant No.2022FY101200).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The source code for MSIMG, along with all scripts used for data processing, statistical analysis, and data visualization in this study, is available on GitHub (<https://github.com/NIM-NMDC/MSIMG>). The public datasets used for the case studies in this research, SPNS and CD, were downloaded from the Metabolomics Workbench database under accession IDs ST001937 and ST003313, respectively.

**Conflicts of Interest:** The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

- Sumner, L.W.; Mendes, P.; Dixon, R.A. Plant metabolomics: Large-scale phytochemistry in the functional genomics era. *Phytochemistry* **2003**, *62*, 817–836. [\[CrossRef\]](#) [\[PubMed\]](#)
- Dettmer, K.; Aronov, P.A.; Hammock, B.D. Mass spectrometry-based metabolomics. *Mass Spectrom. Rev.* **2007**, *26*, 51–78. [\[CrossRef\]](#)
- Bedair, M.; Sumner, L.W. Current and emerging mass-spectrometry technologies for metabolomics. *TrAC Trends Anal. Chem.* **2008**, *27*, 238–250. [\[CrossRef\]](#)
- Lei, Z.; Huhman, D.V.; Sumner, L.W. Mass spectrometry strategies in metabolomics. *J. Biol. Chem.* **2011**, *286*, 25435–25442. [\[CrossRef\]](#) [\[PubMed\]](#)
- Aebersold, R.; Mann, M. Mass spectrometry-based proteomics. *Nature* **2003**, *422*, 198–207. [\[CrossRef\]](#) [\[PubMed\]](#)
- Han, X.; Aslanian, A.; Yates, J.R., III. Mass spectrometry for proteomics. *Curr. Opin. Chem. Biol.* **2008**, *12*, 483–490. [\[CrossRef\]](#)
- Zhou, X.X.; Zeng, W.F.; Chi, H.; Luo, C.; Liu, C.; Zhan, J.; He, S.M.; Zhang, Z. pDeep: Predicting MS/MS spectra of peptides with deep learning. *Anal. Chem.* **2017**, *89*, 12690–12697. [\[CrossRef\]](#)
- Petricoin, E.F.; Liotta, L.A. Mass spectrometry-based diagnostics: The upcoming revolution in disease detection. *Clin. Chem.* **2003**, *49*, 1227–1229. [\[CrossRef\]](#)
- Silva, C.J.; Onisko, B.C.; Dynin, I.; Erickson, M.L.; Requena, J.R.; Carter, J.M. Utility of mass spectrometry in the diagnosis of prion diseases. *Anal. Chem.* **2011**, *83*, 1609–1615. [\[CrossRef\]](#)
- Yoshida, M.; Hatano, N.; Nishiumi, S.; Irino, Y.; Izumi, Y.; Takenawa, T.; Azuma, T. Diagnosis of gastroenterological diseases by metabolome analysis using gas chromatography–mass spectrometry. *J. Gastroenterol.* **2012**, *47*, 9–20. [\[CrossRef\]](#) [\[PubMed\]](#)
- May, J.C.; McLean, J.A. Advanced multidimensional separations in mass spectrometry: Navigating the big data deluge. *Annu. Rev. Anal. Chem.* **2016**, *9*, 387–409. [\[CrossRef\]](#) [\[PubMed\]](#)
- Chaleckis, R.; Meister, I.; Zhang, P.; Wheelock, C.E. Challenges, progress and promises of metabolite annotation for LC–MS-based metabolomics. *Curr. Opin. Biotechnol.* **2019**, *55*, 44–50. [\[CrossRef\]](#) [\[PubMed\]](#)
- Smith, C.A.; Want, E.J.; O’Maille, G.; Abagyan, R.; Siuzdak, G. XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* **2006**, *78*, 779–787. [\[CrossRef\]](#)
- LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [\[CrossRef\]](#)
- Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. Imagenet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [\[CrossRef\]](#)
- Cadow, J.; Manica, M.; Mathis, R.; Reddel, R.R.; Robinson, P.J.; Wild, P.J.; Hains, P.G.; Lucas, N.; Zhong, Q.; Guo, T.; et al. On the feasibility of deep learning applications using raw mass spectrometry data. *Bioinformatics* **2021**, *37*, i245–i253. [\[CrossRef\]](#)
- Huang, J.; Li, Y.; Meng, B.; Zhang, Y.; Wei, Y.; Dai, X.; An, D.; Zhao, Y.; Fang, X. ProteoNet: A CNN-based framework for analyzing proteomics MS-RGB images. *iScience* **2024**, *27*, 111362. [\[CrossRef\]](#)
- Shen, X.; Shao, W.; Wang, C.; Liang, L.; Chen, S.; Zhang, S.; Rusu, M.; Snyder, M.P. Deep learning-based pseudo-mass spectrometry imaging analysis for precision medicine. *Brief. Bioinform.* **2022**, *23*, bbac331. [\[CrossRef\]](#)
- Zhang, F.; Yu, S.; Wu, L.; Zang, Z.; Yi, X.; Zhu, J.; Lu, C.; Sun, P.; Sun, Y.; Selvarajan, S.; et al. Phenotype classification using proteome data in a data-independent acquisition tensor format. *J. Am. Soc. Mass Spectrom.* **2020**, *31*, 2296–2304. [\[CrossRef\]](#)
- Wang, H.; Yin, Y.; Zhu, Z.J. Encoding LC–MS-based untargeted metabolomics data into images toward AI-based clinical diagnosis. *Anal. Chem.* **2023**, *95*, 6533–6541. [\[CrossRef\]](#) [\[PubMed\]](#)
- Song, Y.; Zhang, S.; He, B.; Sha, Q.; Shen, Y.; Yan, T.; Nian, R.; Lendasse, A. Gaussian derivative models and ensemble extreme learning machine for texture image classification. *Neurocomputing* **2018**, *277*, 53–64. [\[CrossRef\]](#)
- He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [\[CrossRef\]](#)
- Najibi, M.; Rastegari, M.; Davis, L.S. G-cnn: An iterative grid based object detector. In Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2369–2377.
- Lempitsky, V.; Zisserman, A. Learning to count objects in images. *Adv. Neural Inf. Process. Syst.* **2010**, *23*, 1324–1332.
- Canny, J. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1986**, *PAMI-8*, 679–698.
- Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
- Bentley, J.L. Multidimensional binary search trees used for associative searching. *Commun. ACM* **1975**, *18*, 509–517. [\[CrossRef\]](#)

28. Zhou, W.; Lin, L.; Jiang, L.Y.; Wu, J.L.; Xu, W.C.; Zhou, Y.; Wang, M.J.; Cao, X.M.; Lin, H.Q.; Yang, J.; et al. Comprehensive plasma metabolomics and lipidomics of benign and malignant solitary pulmonary nodules. *Metabolomics* **2022**, *18*, 71. [[CrossRef](#)] [[PubMed](#)]
29. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
30. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
31. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning. PMLR, Long Beach, CA, USA, 9–15 June 2019; Chaudhuri, K., Salakhutdinov, R., Eds.; 2019; Volume 97, pp. 6105–6114.
32. Alseekh, S.; Aharoni, A.; Brotman, Y.; Contrepois, K.; D’Auria, J.; Ewald, J.; Ewald, J.C.; Fraser, P.D.; Giavalisco, P.; Hall, R.D.; et al. Mass spectrometry-based metabolomics: A guide for annotation, quantification and best reporting practices. *Nat. Methods* **2021**, *18*, 747–756. [[CrossRef](#)]
33. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [[CrossRef](#)]
34. Cai, L.; Zhu, Y. The challenges of data quality and data quality assessment in the big data era. *Data Sci. J.* **2015**, *14*, 2. [[CrossRef](#)]
35. Bengio, Y.; Courville, A.; Vincent, P. Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1798–1828. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.