








## Article

# Well Begun Is Half Done: The Impact of Pre-Processing in MALDI Mass Spectrometry Imaging Analysis Applied to a Case Study of Thyroid Nodules

Giulia Capitoli <sup>1,2,\*</sup> , Kirsten C. J. van Abeelen <sup>3,†</sup> , Isabella Piga <sup>4</sup> , Vincenzo L'Imperio <sup>5</sup> , Marco S. Nobile <sup>6</sup> , Daniela Besozzi <sup>7</sup>  and Stefania Galimberti <sup>1,2</sup> 

- <sup>1</sup> Bicocca Bioinformatics Biostatistics and Bioimaging B4 Center, Department of Medicine and Surgery, University of Milano-Bicocca, 20900 Monza, Italy; stefania.galimberti@unimib.it
  - <sup>2</sup> Biostatistics and Clinical Epidemiology, Fondazione IRCCS San Gerardo Dei Tintori, 20900 Monza, Italy
  - <sup>3</sup> Radboud University Medical Center, Department of Internal Medicine, 6525 AJ Nijmegen, The Netherlands; kirsten.vanabeelen@radboudumc.nl
  - <sup>4</sup> Proteomics and Metabolomics Unit, Department of Medicine and Surgery, University of Milano-Bicocca, 20900 Monza, Italy; isy1987@hotmail.it
  - <sup>5</sup> Pathology Unit, Fondazione IRCCS San Gerardo dei Tintori, Department of Medicine and Surgery, University of Milano-Bicocca, 20900 Monza, Italy; vincenzo.limperio@unimib.it
  - <sup>6</sup> Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, 30100 Venice, Italy; marco.nobile@unive.it
  - <sup>7</sup> Department of Informatics, Systems, and Communication, University of Milano-Bicocca, 20126 Milan, Italy; daniela.besozzi@unimib.it
- \* Correspondence: giulia.capitoli@unimib.it; Tel.: +39-0264488333  
† These authors contributed equally to this work.

## Abstract

The discovery of proteomic biomarkers in cancer research can be effectively performed in situ by exploiting Matrix-Assisted Laser Desorption Ionization (MALDI) Mass Spectrometry Imaging (MSI). However, due to experimental limitations, the spectra extracted by MALDI-MSI can be noisy, so pre-processing steps are generally needed to reduce the instrumental and analytical variability. Thus far, the importance and the effect of standard pre-processing methods, as well as their combinations and parameter settings, have not been extensively investigated in proteomics applications. In this work, we present a systematic study of 15 combinations of pre-processing steps—including baseline, smoothing, normalization, and peak alignment—for a real-data classification task on MALDI-MSI data measured from fine-needle aspirates biopsies of thyroid nodules. The influence of each combination was assessed by analyzing the feature extraction, pixel-by-pixel classification probabilities, and LASSO classification performance. Our results highlight the necessity of fine-tuning a pre-processing pipeline, especially for the reliable transfer of molecular diagnostic signatures in clinical practice. We outline some recommendations on the selection of pre-processing steps, together with filter levels and alignment methods, according to the mass-to-charge range and heterogeneity of data.

**Keywords:** pre-processing; MALDI; mass spectrometry; machine learning; feature design; classification performance; thyroid nodules



Academic Editor: Wei Zhu

Received: 19 March 2025

Revised: 17 June 2025

Accepted: 27 June 2025

Published: 10 July 2025

**Citation:** Capitoli, G.; van Abeelen, K.C.J.; Piga, I.; L'Imperio, V.; Nobile, M.S.; Besozzi, D.; Galimberti, S. Well Begun Is Half Done: The Impact of Pre-Processing in MALDI Mass Spectrometry Imaging Analysis Applied to a Case Study of Thyroid Nodules. *Stats* **2025**, *8*, 57. <https://doi.org/10.3390/stats8030057>

**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

Over the last ten years, Matrix-Assisted Laser Desorption Ionization (MALDI) Mass Spectrometry Imaging (MSI) has emerged as one of the key technologies for cancer

biomarker discovery directly in situ [1,2]. The feasibility of identifying molecular profiles and building classification models based on MALDI-MSI has been demonstrated in several cancer-related contexts [3–5], including pancreatic adenocarcinoma [6], thyroid cancer [7], and skin disease [8]. These studies showed that MALDI-MSI can support histopathological diagnosis by capturing molecular heterogeneity directly in tissue sections. The technology visualizes the distribution of molecules in a biological sample [1] in order to differentiate remarkable spectra profiles, namely, mass-to-charge ( $m/z$ ) ratios and corresponding intensities. However, due to experimental and instrumental limitations (including ion suppression effects, matrix inconsistencies, and spot-to-spot variability), the spectra extracted by MALDI-MSI can be noisy. This variability can significantly affect the reliability of classification models and biomarker discovery. Studies exploiting MALDI-MSI data therefore often perform various pre-processing steps to reduce the intrinsic instrumental and analytical variability. The standard pre-processing procedure consists of four steps—baselining, smoothing, normalization, and alignment [9]—that are followed by the two consolidated stages of peak detection and quantification. However, pre-processing has been under-reported in the mass spectrometry literature, even when performed, and it has received limited methodological attention. Previous studies have evaluated individual preprocessing techniques such as baselining, denoising, and normalization [10,11], or proposed automated pipelines that select preprocessing steps via heuristic methods [12,13]. However, the impact of different combinations of preprocessing methods, and their order of application, has received far less attention, particularly in the MALDI-MSI domain. Related research on SELDI-MS data, a variant of MALDI-MSI, has shown that such combinations can drastically influence peak detection and quantification, as well as classification outcomes [14–16]. The importance and the effect of each separate pre-processing step, as well as their combination, have not yet been extensively investigated for MALDI-MSI data [17–20]. This underscores the importance of evaluating not just individual steps but also their sequence of application.

More recently, user-friendly platforms such as MALDI<sub>m</sub>ID [21], MSiReader [22], MALDIViz [23], and Cardinal [24] have been developed to facilitate mass spectrometry imaging data analysis. These tools are widely used and provide intuitive graphical user interfaces; however, they either require users to input already pre-processed data or offer broad flexibility in selecting pre-processing steps, without clear guidance or standardization. While this flexibility is appropriate for exploratory research, it also shifts the responsibility—and the variability—onto the user, often without assessing how these decisions affect downstream analyses. These considerations become particularly relevant when the objective is to build diagnostic or predictive models intended for clinical application. In such cases, reproducibility and robustness are essential, and the preprocessing pipeline becomes a critical component of the analytical workflow. Therefore, shedding light on how preprocessing choices influence classification performance is a necessary step toward more transparent, standardized, and clinically translatable MALDI-MSI analysis.

The main objective of this study was to systematically evaluate how different pre-processing strategies affect the molecular feature space and classification performance of MALDI-MSI data, using a real-world dataset of thyroid fine-needle aspirates. Unlike previous works, we focused on the joint effect of preprocessing combinations and filtering parameters, providing practical insight for the preprocessing phase in MSI-based diagnostic workflows. This study broadens previous works [14,25] and fills a gap by analyzing additional pre-processing combinations for an MALDI-MSI dataset. Namely, 15 different pre-processing combinations under three different scenarios regarding inter- and intra-patient filter parameters, as well as two alignment settings, were explored. We show that the way MALDI-MSI spectra are pre-processed affects the performance of the designed

classifier in terms of molecular signature and classification outcomes. As a case in point, we used real data from MALDI-MSI analysis of thyroid fine-needle aspirates (FNA) biopsies [26]. Our results highlight the importance of the pre-processing stage in research, since it can heavily influence subsequent analysis; indeed, in our use-case, the success of classification depended on the pre-processed data received. We conclude the paper by outlining some recommendations on the selection of pre-processing steps, filter levels, and alignment methods, based on the overall findings of our study. The results demonstrate that inadequate preprocessing may lead not only to decreased accuracy, but also to incorrect classifications. Our work thus contributes to closing a critical methodological gap in the use of MALDI-MSI for diagnostic purposes by evidencing the non-trivial impact of preprocessing decisions on statistical analysis.

## 2. Materials and Methods

### 2.1. Data Pre-Processing

The spectra composed of mass-to-charge ( $m/z$ ) values were pre-processed using one or more of the following steps (see the Appendix A for details):

- (B) Baseline subtraction: electrical noise and chemical impurities in the sample were estimated and subtracted from the spectrum.
- (S) Smoothing: interfering peaks from sources unrelated to the patient's sample were removed.
- (N) Normalization: spectra were brought to the same intensity range.
- (A) Alignment: slight differences in  $m/z$ -values were corrected, so that the same proteins could be identified between spectra.
- (P) Peak detection: in each pre-processed spectrum, peaks were extracted as features and their  $m/z$ -values were aligned.

These steps were followed by intra- and inter-patient filtering, to retain only the features shared by at least a certain percentage of spectra within the same samples, or among samples with the same diagnosis. The standard approaches used in the pre-processing steps were as follows: median method for baseline subtraction; moving average (MA) method for smoothing; total ion current (TIC) method for normalization; and mean absolute deviation (MAD) noise estimation method for alignment, as well as for peak detection. In addition, different values for the parameter of the intra- and inter-patient filters were analyzed, with filter values equal to 5% (Scenario I), 25% (Scenario II), and 50% (Scenario III). Further details can be found in Table A1. Finally, the influence of joint alignment with respect to individual alignment was investigated: in the first case, the spectra of different samples were aligned separately; in the second case, the spectra of all patients were aligned together.

### 2.2. Combinations of Pre-Processing Steps and Parameter Settings

We assumed a strict ordering of the four pre-processing steps based on baselining, smoothing, normalization, and alignment (e.g., B must be performed before S, N, and A). This resulted in exactly 15 possible combinations (Table 1). Peak detection was applied after each combination in order to extract an intensity matrix for further analysis.

### 2.3. Case Study and Statistical Analysis

The proteomics data we used originated from a study aiming to investigate whether the application of MALDI-MSI on FNAs of thyroid nodules could enhance correct classification of Hashimoto thyroiditis (HT), hyperplastic nodule (Hp), and papillary thyroid carcinoma (PTC) [26]. Data were acquired in positive-ion mode, in the  $m/z$  range 3000–15,000 Da, with a laser focus setting of 50  $\mu\text{m}$  and a pixel size of 50  $\times$  50  $\mu\text{m}$  with an UltrafleXtreme mass spectrometer (Bruker Daltonics, Bremen, Germany). They were exported

in imzML format (pixel-by-pixel MALDI-MSI analysis) and subsequent pre-processing was performed using the MALDIquant R package. More details on the study protocol, in particular on the technical specifics of in situ proteomics MALDI-MSI analysis and its results, have been previously described [26].

**Table 1.** Pre-processing combinations.

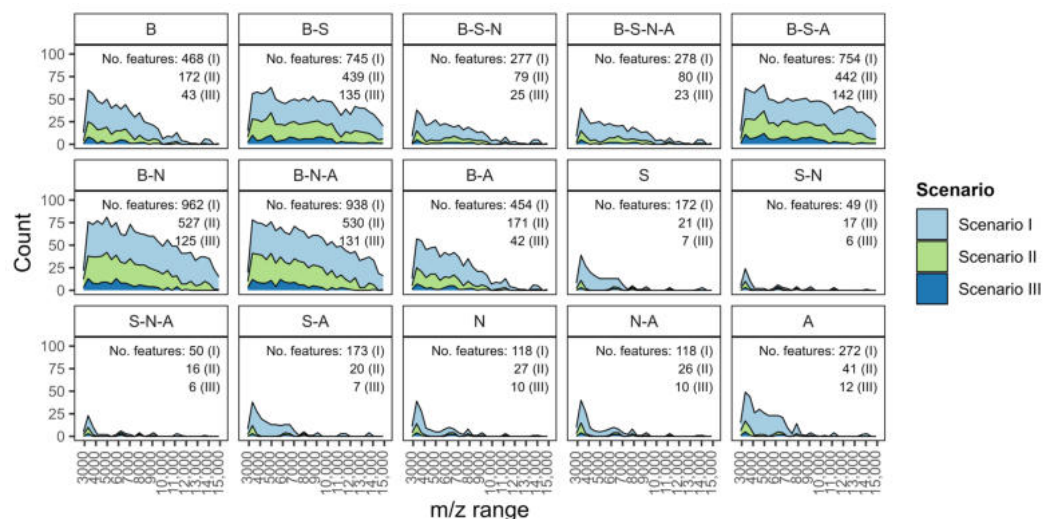
Combination	Baseline	Smoothing	Normalization	Alignment	Peak Detection
B	X				X
B-S	X	X			X
B-S-N	X	X	X		X
B-S-N-A	X	X	X	X	X
B-S-A	X	X		X	X
B-N	X		X		X
B-N-A	X		X	X	X
B-A	X			X	X
S		X			X
S-N		X	X		X
S-N-A		X	X	X	X
S-A		X		X	X
N			X		X
N-A			X	X	X
A				X	X

Briefly, a cohort of 27 patients (9 Hp, 9 HT, and 9 PTC) was used for training. A total of 126 annotated pathological areas referred to as regions of interest (ROIs), containing informative pixels from the MSI data, were employed to train the classification model. A multinomial least absolute shrinkage and selection operator (LASSO) model was applied to identify discriminating  $m/z$ -values between malignant (PTC) and benign (Hp and HT) thyroid cells. For model validation, an independent cohort of 13 patients (2 Hp, 3 HT, and 8 PTC) was used to evaluate the classification performance and to assess the influence of different pre-processing pipelines on predictions at the pixel level. The classifier predicted a probability for the three diagnostic classes for each spectrum, corresponding to a pixel in the MSI dataset of the validation cohort. The highest of these three probabilities determined the label of the pixel. A patient was then classified as malignant when the ratio of malignant pixels vs. the total number of pixels was greater than a specified threshold, or conversely benign. This threshold was identified by applying the Youden index, maximizing sensitivity and specificity. For performance evaluation, the three diagnostic classes were reduced to two (benign vs. malignant), in line with previous work [26], to allow for quantitative assessment of the effect of pre-processing on classification accuracy. All the analyses were performed using the open-source R software v.4.1.1 (R Foundation for Statistical Computing, Vienna, Austria).

### 3. Results

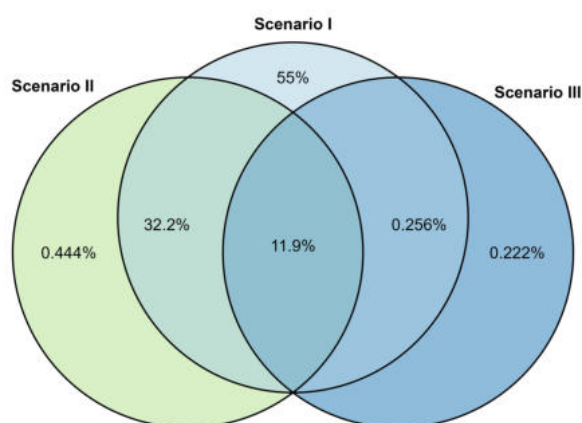
#### 3.1. Feature Extraction from Pre-Processing

The impact of the 15 pre-processing combinations described in Table 1 and of the intra- and inter-patient filters on the amount of extracted features, as well as their localization over the  $m/z$  axis, is shown in Figure 1.



**Figure 1.** Number of features ( $m/z$ -values) extracted by the 15 pre-processing combinations in each scenario (Scenario I—5% filter, II—25% filter, and III—50% filter). The total number of included signals in each setting is indicated in the upper right corner.

Notably, combinations starting with smoothing (S-), normalization (N-), or alignment (A-) tended to result in features in the range between 3000 and 8000  $m/z$  values. In contrast, combinations starting with the baseline (B-) included features across the whole  $m/z$  range. This behavior can be attributed to the exponential baseline often observed in spectra of intact proteins, particularly in the lower  $m/z$  region. Without early baseline correction, subsequent steps such as normalization may introduce bias, leading to the underrepresentation of certain spectral regions. Moreover, there was a clear gradient in the impact of the different filter levels, with the lower filter (Scenario I) systematically dominating over the  $m/z$  range. A summary of the overlapping  $m/z$  features over the three scenarios is visualized in Figure 2, where it can be noted that approximately 12% of the extracted features were shared among the three scenarios, and approximately 32% of the signals were in common between Scenario I and II.

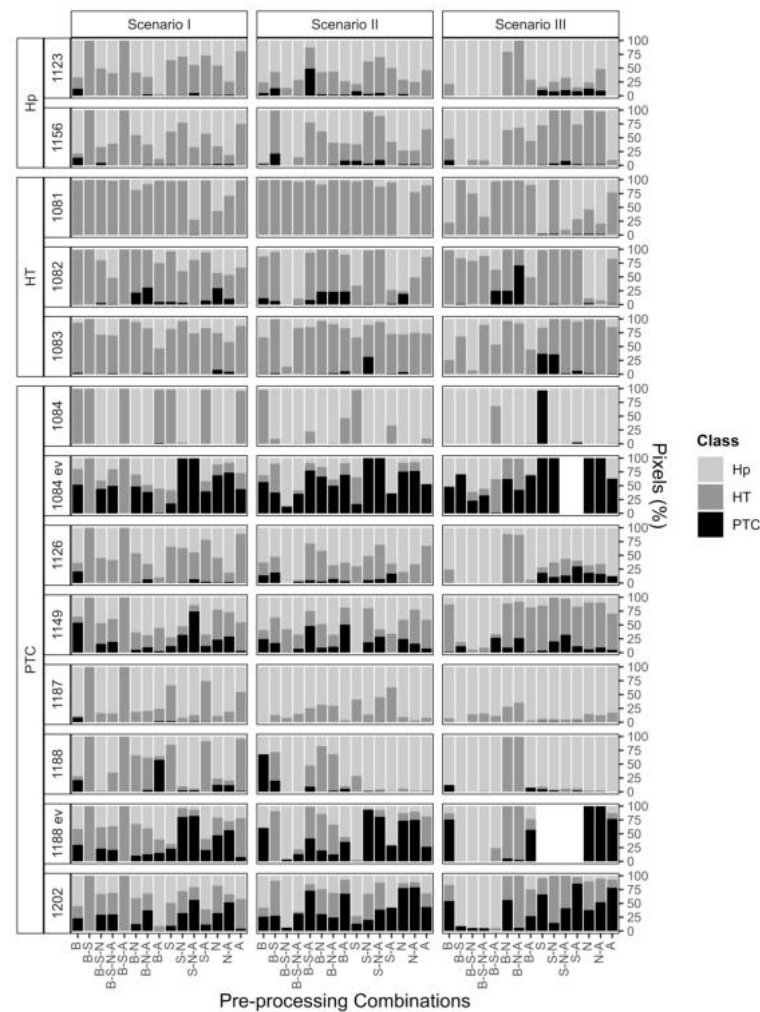


**Figure 2.** Venn diagram visualizing the overlap of selected  $m/z$  values between Scenario I (5% filter), II (25% filter), and III (50% filter).

Concerning the joint and individual alignment, they turned out to be mainly similar except for the combination B-S-A, where a larger number of included features can be observed for individual alignment (Figure A1). The influence that each patient of the validation cohort had on feature extraction is shown in Figure A2.

### 3.2. Pixel-by-Pixel Classification

The classification probabilities of being Hp, HT, and PTC in the pixel-by-pixel approach for the validation cohort are visualized in Figure 3.

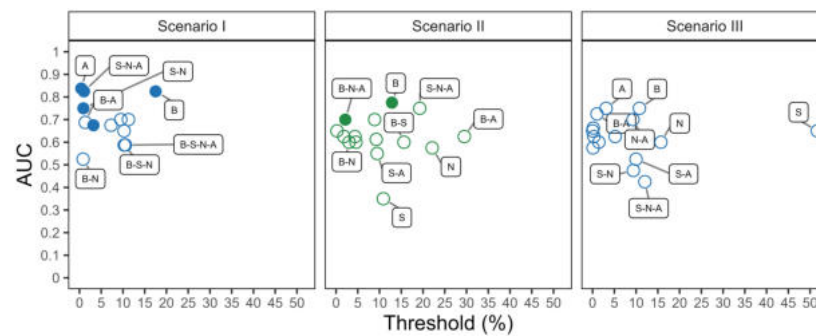


**Figure 3.** Distributions of the diagnostic predicted probabilities per validation patient by pre-processing combinations and scenarios. The stacked bars indicate the percentage of single pixels classified as Hp, HT, and PTC for each setting.

A marked heterogeneity in the distribution of the three predicted diagnoses was present among the scenarios, as well as among pre-processing combinations within each scenario. Overall, the benign cases were well predicted, even if in some situations the specific diagnostic subtype was not recognized. On the other hand, the signals of malignancies were only clearly visible in samples rich in cellularity (Table A2) and in Scenario II with the 25% filter. Reducing the intra- and inter-filter value from 25% to 5% (Scenario I) had a direct impact, with predictions shifting in favor of the benign classes (e.g., patient 1202). Lastly, increasing the intra- and inter-filter value from 25% to 50% (Scenario III) produced pronounced changes in the predictions (e.g., patient 1188 ex vivo). The intra-patient filter removed noisy features inconsistently expressed within a single sample, while the inter-patient filter eliminated signals not consistently shared among patients with the same diagnosis. While these filters helped reduce technical and biological variability, excessive filtering may lead to the exclusion of diagnostically relevant features, especially in heterogeneous samples such as FNAs. Our findings emphasize the need to carefully calibrate these parameters based on data quality, biological complexity, and diagnostic goals.

### 3.3. Classification Performance

The AUC, sensitivity, specificity, and optimal threshold calculated for all the combinations in each scenario are summarized in Figure 4 and in Tables A3–A5. Comparison among the scenarios shows that reducing the intra- and inter-patient filter from 25% to 5% positively affected the results, with increased AUC and sensitivity for the majority of combinations (e.g., B, B-A, S-N-A, and N), but decreased the threshold of malignancy. In addition, increasing the intra- and inter-patient filter from 25% to 50% (Scenario III) yielded a lower threshold and classification performance.



**Figure 4.** Summary of AUC and thresholds by pre-processing combinations and scenarios. Combinations with sensitivity  $\geq 0.75$  and specificity  $\geq 0.80$  are shown as filled circles (otherwise as empty circles) at the optimal threshold.

In terms of classification metrics, the highest AUC in Scenario I (5% filter) was obtained with the A combination (AUC = 0.84, sensitivity = 0.75, specificity = 1.00), followed closely by the B and S-N-A combinations (AUC = 0.83). These combinations consistently produced high sensitivity and specificity at their optimal thresholds. In Scenario II (25% filter), the best performing combination was again B (AUC = 0.78, sensitivity = 0.75, specificity = 1.00), with S-N-A and B-S-A also showing acceptable performance (AUC = 0.75 and 0.70, respectively). In Scenario III (50% filter), the top-performing combinations were A and B (AUC = 0.75), but with reduced sensitivity compared to Scenario I (AUC = 0.63 and AUC = 0.50, respectively). Notably, the performance of S-N-A decreased significantly in Scenario III (AUC = 0.43), highlighting the interaction between filter settings and the effectiveness of specific pre-processing chains. These results confirm that both the choice of pre-processing steps and the associated filtering thresholds critically impacted the classifier’s ability to separate benign from malignant nodules, and that some combinations may only perform well under specific parameter settings.

It is, however, important to emphasize that the optimal decision thresholds vary significantly across scenarios and pre-processing combinations. While a perfect classification performance can be achieved under certain settings, clinical interpretation remains essential. A high threshold—meaning that a high percentage of pixels classified as malignant are required to label a sample as malignant—may increase the risk of false negatives, especially in heterogeneous or low-cellularity samples. Conversely, a low threshold may lead to false positives, particularly in noisy or atypical yet benign cases, where a small proportion of pixels classified as malignant could result in a misdiagnosis. Given that a classifier operates in a three-class setting and assigns the class with the highest probability, a pixel with just over 33% probability of malignancy is already labeled as malignant. Therefore, if a sample is classified as malignant based on a global threshold of less than 1% malignant pixels, there is a substantial risk of overdiagnosis. To address this issue, in our previous study [7], we introduced two distinct thresholds—one for “certain benignity” and one for “certain malignancy”—as well as an intermediate gray zone designated for close follow-up, aiming to improve clinical accuracy and reduce diagnostic uncertainty.

## 4. Discussion

In this work, we have shown that the sequence of pre-processing steps for the analysis of MALDI-MSI data and the parameter setup might have a large impact on the feature extraction, classification predictions, and classification performance for a molecular diagnostic signature.

Both the total number of features included in the classifier and the  $m/z$  range of these features changed significantly depending on how pre-processing was performed. Lowering or increasing the intra- and inter-patient filter decreased or increased the number of features across the entire  $m/z$  range. A higher filter value may be less suitable since valuable information may be excluded from the model, which may be the case in heterogeneous datasets as used in this study. The lower filter values allow the model to include this information, improving classification performance. Furthermore, some pre-processing combinations retained the entire  $m/z$  ranges, which may be suitable for clinical studies where no prior information on peaks of interest is available. Even if retaining a high amount of  $m/z$  values may include a substantial amount of noise, this can be controlled with intra- and inter-patient filters. Moreover, the use of joint instead of individual alignment in the training and validation samples also influenced feature selection, where the included  $m/z$  features in the model could drastically change.

Furthermore, a comparison between pre-processing combinations of the pixel-by-pixel classification probabilities for the three considered diagnostic classes showed significantly varying predictions across the combinations. The analysis revealed the difficulty of identifying malignancy in samples with a poor cellular composition for the majority of the pre-processing combinations, where the presence of inflammation, a poor background, or a low percentage of thyroicities in the sample lead to an incorrect benignity prediction by the classifier.

Lastly, the analysis of pre-processing combinations through the classification performance allowed for a direct comparison of the AUC index. It was shown that the lower filter of 5% positively impacted the performance for the majority of combinations, but a wide range of thresholds for malignancy were obtained with high AUC values. Notably, the low threshold may be not suitable, since this increases the possibility of false positives due to technical artifacts.

Previous research [16,20] evaluated pre-processing procedures consisting of predefined steps for SELDI-MS that were only partially similar to the ones we considered here, with findings indicating that factors other than the pre-processing method, e.g., parameter settings, can influence the subsequent analysis. The two parameters considered in this study, the intra- and inter-patient filter and the alignment option, were also shown to impact classification and feature selection. As noted in previous research [20], extending the examination to a wider range of parameter options for each step is recommended, although this opens up a very extensive assessment aiming to optimize the parameter choice, which will be the subject of a future study.

While we used proteomics data from thyroid FNA biopsies to investigate the role of pre-processing on the results of statistical analyses, we believe that the heterogeneity of these findings can be, in principle, generalized to other clinical contexts, and that there are no absolute rules for all situations. Therefore, based on the overall results of this study, the following recommendations are outlined:

1. **Pre-processing combinations.** Select the pre-processing steps with care, as they directly influence not only the number and nature of extracted features, but also their distribution across the  $m/z$  range. For example, combinations beginning with baseline correction tend to preserve features across a wider  $m/z$  interval, whereas others may prioritize narrower but potentially more specific spectral zones. Understanding how

different combinations behave for your dataset is essential to ensure that relevant biological signals are retained and emphasized in downstream analysis.

2. **Intra- and inter-patient filter setting.** The filter's value has an impact on the overall number of extracted features included in the model for signature identification. Low filter levels may better fit with heterogeneous data, while high filter levels may better fit with homogeneous data. Importantly, intra- and inter-patient filtering do not indiscriminately reduce the feature space, but rather reduce unwanted variability: intra-patient filters help control for artifacts and noise within a sample, while inter-patient filters (applied among samples with the same diagnostic label) reduce variability not associated with the disease itself. Nonetheless, excessive filtering can be harmful, especially in heterogeneous cohorts. Therefore, such parameters must be carefully evaluated according to the context. Our results indicate that lower filtering levels often better preserve diagnostically relevant information.
3. **Individual over joint alignment.** Individual alignment is favored as it ensures that the inclusion of new patients in the validation cohort does not retroactively affect the spectra of the training cohort and the associated feature set. This is particularly relevant when aiming for robust and reproducible workflows in prospective studies. In our analysis, individual alignment provided more consistent and reliable feature sets across patients, whereas joint alignment introduced variations that could confound model interpretation and performance.

We do not propose a universally optimal pre-processing combination, as this depends on the experimental conditions and downstream objectives. Our findings may serve as a starting point for similar diagnostic applications, with the understanding that they should be further evaluated in each new context. Reporting the pre-processing workflow in research studies should be mandatory, as it can improve the reproducibility of the research performed and allow for a better understanding of this essential procedure by the scientific community.

## 5. Conclusions

Pre-processing is of paramount importance in the management of mass spectrometry imaging (MSI) data, yet it remains an under-documented phase in many biomedical studies. Our findings demonstrate that pre-processing is not a trivial step, but one that significantly impacts the accuracy, robustness, and interpretability of downstream classification models. In this study, we systematically evaluated 15 preprocessing combinations and multiple filtering scenarios on a real-world MALDI-MSI dataset of thyroid fine-needle biopsies. We demonstrated that improper or suboptimal preprocessing pipelines can result in poor classification performance, potentially leading to misclassifications between benign and malignant nodules. These results underscore the necessity of carefully tuning and reporting all pre-processing steps when MSI data are used for clinical diagnostics or predictive modeling. Furthermore, we emphasize the importance of selecting preprocessing strategies that are context-aware by considering tissue heterogeneity, cellular composition, and the final application (exploratory versus clinical). The implications of this work go beyond thyroid pathology: the outlined methodology and evidence may serve as a generalizable framework for evaluating the effect of preprocessing in other proteomic and imaging contexts. By making the pipelines more transparent and evidence-driven, this research can contribute to the development of reproducible and clinically viable MALDI-MSI workflows. Future work could consider exploring the parameter options and settings within each pre-processing step and incorporating additional classification models to validate the generalizability of these observations.

**Author Contributions:** Conceptualization, G.C.; methodology, G.C. and K.C.J.v.A.; software, G.C. and K.C.J.v.A.; validation, G.C. and K.C.J.v.A.; formal analysis, K.C.J.v.A.; investigation, G.C. and K.C.J.v.A.; resources, I.P. and V.L.; data curation, I.P. and V.L.; writing—original draft preparation, G.C. and K.C.J.v.A.; writing—review and editing, I.P., V.L., M.S.N., D.B., and S.G.; visualization, G.C. and K.C.J.v.A.; supervision, M.S.N., D.B., and S.G.; project administration, V.L. and S.G.; funding acquisition, D.B. and S.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Ricerca Finalizzata GR-2019-12368592, by the Italian Ministry of University MUR Dipartimenti di Eccellenza 2023-2027 (l. 232/2016, art. 1, commi 314–337) and by the National Plan for NRRP Complementary Investments (established with the decree-law May 6, 2021, n. 59, converted by law n. 101 of 2021) in the call for the funding of research initiatives for technologies and innovative trajectories in the health care sectors (Directorial Decree n. 931 of 06-06-2022)—project n. PNC0000003—AdvaNced Technologies for Human-centrEd Medicine (project acronym: ANTHEM). This work reflects only the authors' views and opinions, neither the Ministry for University and Research nor the European Commission can be considered responsible for them.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of ASST Monza HSG (protocol code 18445 and date of approval 27 December 2016).

**Informed Consent Statement:** The study was carried out in accordance with the relevant guidelines and regulations. It was approved by the ASST Monza Ethical Board (Associazione Italiana Ricerca sul Cancro- AIRC-MFAG 2016 Id. 18445, HSG Ethical Board Committee approval October 2016, 27 December 2016). Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Data that support the findings of this study are available on request from the corresponding author G.C., upon reasonable request.

**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.

## Abbreviations

The following abbreviations are used in this manuscript:

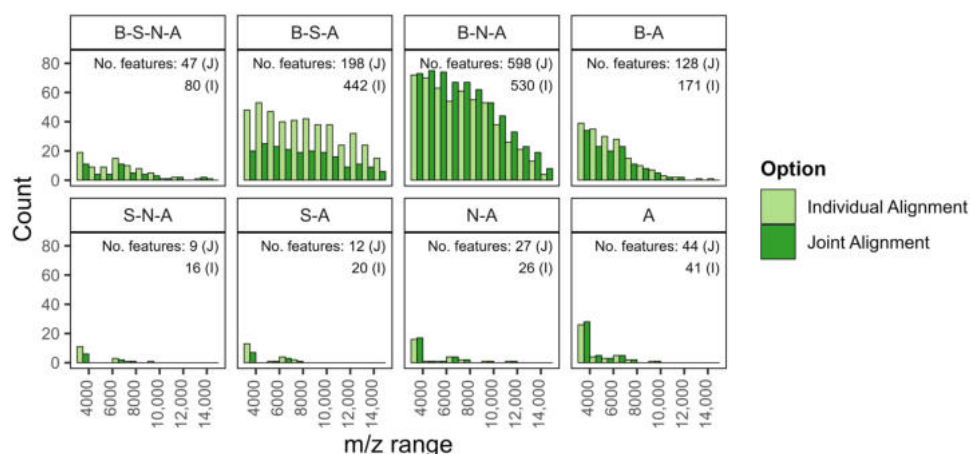
AUC	Area Under Curve
FNA	Fine Needle Aspiration
HP	Hyperplastic
HT	Hashimoto Thyroiditis
LASSO	Least Absolute Shrinkage and Selection Operator
MALDI	Matrix Assisted Laser Desorption Ionization
MSI	Mass Spectrometry Imaging
PTC	Papillary Thyroid Carcinoma
ROC	Receiver Operating characteristic Curve

## Appendix A. Additional Information, Tables & Figures

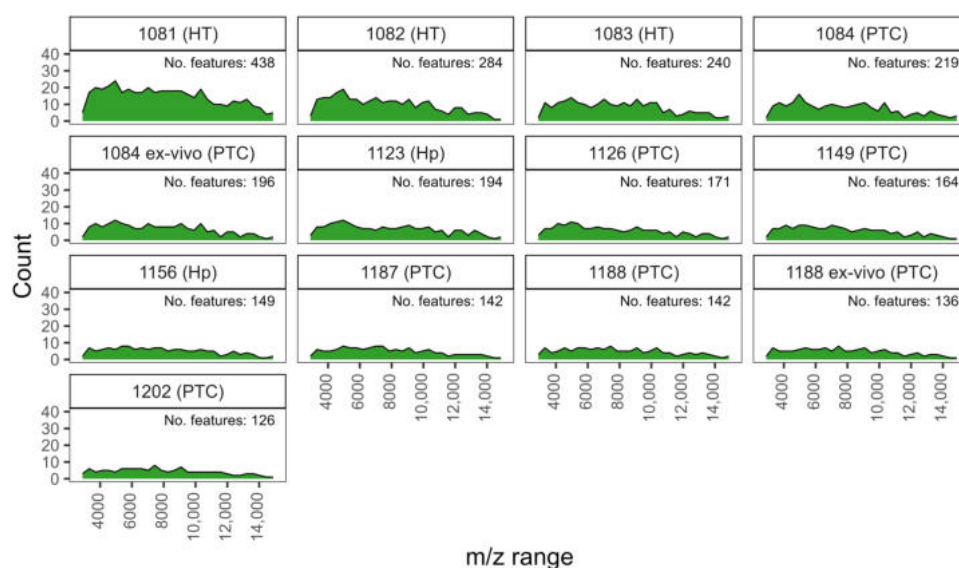
### Appendix A.1. Computational Resources

The analyses were performed using High Performance Computing (HPC) system Marconi100. Data preprocessing (MALDIquant v.1.21), classification (glmnet v.4.1-1), and performance analysis (pROC v.1.18.0) were performed using the open-source R software v.4.0.1. Visualizations were created with the help of plyr v.1.0.9, reshape v.0.8.9, writexl v.1.4.0, extrafont v.0.18, VennDiagram v.1.7.3, ggplot2 v.3.3.6, and additions ggh4x v.0.2.1 and ggrepel v.0.9.1. The 95% confidence intervals for the AUC scores were determined through the DeLong method, while the 95% confidence intervals for sensitivity and specificity for a given threshold were determined through stratified bootstrapping with 2000 replicates.

Appendix A.2. Figures



**Figure A1.** Histograms of  $m/z$  values selected by the classification model for Scenario II (25% intra- and inter- patient filter setting), with the option of individual or joint alignment within all tested pre-processing combinations that included the alignment step (B: Baseline, S: Smoothing, N: Normalization, and A: Alignment). The total number of included features for both options are indicated in the upper right corner for each combination. For joint alignment, the included features are averaged over the validation cohort, as an exclusive classification model was created per validation patient.



**Figure A2.** An illustrative case:  $m/z$  values selected by the classification models (one per patient) for the pre-processing combination B-S-A (Baseline, Smoothing, Alignment), where joint alignment was used. The total number of included features is indicated in the upper right corner for each patient.

Appendix A.3. Tables

**Table A1.** Details on pre-processing steps.

Name	Step	Method	Parameters
B	Baseline	median method	
S	Smoothing	moving average	window equal to five
N	Normalization	total ion current	
A	Alignment	mean absolute deviation	half window width of five
P	Peak detection	mean absolute deviation	half window equal to five signal-to-noise ratio equal to six

**Table A2.** Clinical and histological information for the validation cohort. The quality of the cellular composition of the sample is based on observations by the pathologist.

Patient ID	Cellular Composition	Diagnosis	
		Cytological	Histological
1081	Rich	TIR2	HT
1082	Good	TIR3	HT
1083	Rich	TIR2	HT
1084	Poor	TIR5	PTC
1084 ex-vivo	Rich	TIR5	PTC
1123	Good	TIR2	Hp
1126	Poor	TIR5	PTC
1149	Good	TIR5	PTC
1156	Good	TIR2	Hp
1187	Poor	TIR5	PTC
1188	Poor	TIR5	PTC
1188 ex-vivo	Good	TIR5	PTC
1202	Rich	TIR4	PTC

**Table A3.** Results from the ROC-AUC analysis for Scenario I (5% intra- and inter- patient filter setting), sorted from best to lowest AUC score. The 95% confidence intervals are denoted in parentheses.

Combination	AUC	Classification Performance		
		Threshold	Sensitivity	Specificity
A	0.84 (0.60, 1.00)	0.4%	0.75 (0.50, 1.00)	1.00 (1.00, 1.00)
B	0.83 (0.57, 1.00)	18%	0.75 (0.50, 1.00)	1.00 (1.00, 1.00)
S-N-A	0.83 (0.57, 1.00)	1%	0.88 (0.63, 1.00)	0.80 (0.40, 1.00)
S-N	0.75 (0.46, 1.00)	9%	0.75 (0.38, 1.00)	0.80 (0.40, 1.00)
S-A	0.70 (0.39, 1.00)	10%	0.50 (0.13, 0.88)	1.00 (1.00, 1.00)
N-A	0.70 (0.39, 1.00)	11%	0.63 (0.25, 0.88)	1.00 (1.00, 1.00)
B-A	0.69 (0.37, 1.00)	1%	0.63 (0.25, 0.88)	0.80 (0.40, 1.00)
B-N-A	0.68 (0.34, 1.00)	3%	0.75 (0.50, 1.00)	0.80 (0.40, 1.00)
S	0.68 (0.36, 0.99)	7%	0.50 (0.13, 0.88)	1.00 (1.00, 1.00)
N	0.65 (0.32, 0.98)	10%	0.63 (0.25, 0.88)	0.80 (0.40, 1.00)
B-S-N	0.59 (0.25, 0.92)	10%	0.50 (0.13, 0.88)	1.00 (1.00, 1.00)
B-S-N-A	0.59 (0.25, 0.92)	11%	0.63 (0.25, 1.00)	1.00 (1.00, 1.00)
B-S	0.50 (0.50, 0.50)		0 or 1	1 or 0
B-S-A	0.50 (0.50, 0.50)		0 or 1	1 or 0

**Table A4.** Results from the ROC-AUC analysis for Scenario II (25% intra- and inter- patient filter setting), sorted from best to lowest AUC score. The 95% confidence intervals are denoted in parentheses.

Combination	AUC	Classification Performance		
		Threshold	Sensitivity	Specificity
B	0.78 (0.48, 1.00)	13%	0.75 (0.38, 1.00)	1.00 (1.00, 1.00)
S-N-A	0.75 (0.46, 1.00)	19%	0.50 (0.13, 0.88)	1.00 (1.00, 1.00)
B-S-A	0.70 (0.38, 1.00)	9%	0.63 (0.25, 0.88)	0.80 (0.40, 1.00)
B-N-A	0.70 (0.37, 1.00)	2%	0.75 (0.38, 1.00)	0.80 (0.40, 1.00)
B-S-N	0.65 (0.36, 0.94)	0.2%	0.50 (0.13, 0.88)	1.00 (1.00, 1.00)
B-S-N-A	0.63 (0.27, 0.98)	2%	0.63 (0.25, 1.00)	1.00 (1.00, 1.00)
B-A	0.63 (0.30, 0.95)	30%	0.50 (0.13, 0.88)	1.00 (1.00, 1.00)
A	0.63 (0.30, 0.95)	4%	0.50 (0.13, 0.88)	1.00 (1.00, 1.00)
N-A	0.61 (0.28, 0.94)	9%	0.50 (0.13, 0.88)	1.00 (1.00, 1.00)
B-S	0.60 (0.26, 0.94)	16%	0.63 (0.25, 0.88)	0.80 (0.40, 1.00)
B-N	0.60 (0.26, 0.94)	3%	0.63 (0.25, 1.00)	0.80 (0.40, 1.00)
S-N	0.60 (0.26, 0.94)	5%	0.63 (0.25, 1.00)	0.80 (0.40, 1.00)
N	0.58 (0.23, 0.92)	22%	0.50 (0.13, 0.88)	1.00 (1.00, 1.00)
S-A	0.55 (0.20, 0.90)	9%	0.50 (0.13, 0.88)	1.00 (1.00, 1.00)
S	0.35 (0.02, 0.68)	11%	0.25 (0.00, 0.50)	1.00 (1.00, 1.00)

**Table A5.** Results from the ROC-AUC analysis for Scenario III (50% intra- and inter- patient filter setting), sorted from best to lowest AUC score. The 95% confidence intervals are denoted in parentheses.

Combination	AUC	Classification Performance		
		Threshold	Sensitivity	Specificity
B	0.75 (0.48, 1.00)	11%	0.50 (0.13, 0.88)	1.00 (1.00, 1.00)
A	0.75 (0.46, 1.00)	3%	0.63 (0.25, 0.88)	1.00 (1.00, 1.00)
B-A	0.73 (0.43, 1.00)	1%	0.63 (0.25, 0.88)	1.00 (1.00, 1.00)
N-A	0.70 (0.38, 1.00)	9%	0.63 (0.25, 0.88)	1.00 (1.00, 1.00)
B-S	0.63 (0.38, 0.88)	5%	0.38 (0.00, 0.75)	1.00 (1.00, 1.00)
B-S-N-A	0.63 (0.38, 0.87)	0.3%	0.38 (0.13, 0.88)	0.80 (0.40, 1.00)
B-S-N	0.65 (0.38, 0.92)	0.004%	0.50 (0.125, 0.875)	0.80 (0.40, 1.00)
B-N	0.66 (0.34, 0.97)	0.2%	0.63 (0.25, 0.88)	0.80 (0.40, 1.00)
S	0.65 (0.33, 0.97)	52%	0.38 (0.00, 0.75)	1.00 (1.00, 1.00)
B-S-A	0.58 (0.29, 0.86)	0.09%	0.38 (0.00, 0.75)	0.80 (0.40, 1.00)
B-N-A	0.60 (0.27, 0.93)	1%	0.50 (0.13, 0.88)	0.80 (0.40, 1.00)
N	0.60 (0.25, 0.95)	16	0.50% (0.13, 0.88)	0.80 (0.40, 1.00)
S-A	0.53 (0.19, 0.86)	0%	0.38 (0.00, 0.75)	1.00 (1.00, 1.00)
S-N	0.48 (0.13, 0.82)	9%	0.50 (0.13, 0.88)	0.80 (0.40, 1.00)
S-N-A	0.43 (0.08, 0.77)	12%	0.38 (0.13, 0.75)	1.00 (1.00, 1.00)

## References

- Rohner, T.C.; Staab, D.; Stoeckli, M. MALDI mass spectrometric imaging of biological tissue sections. *Mech. Ageing Dev.* **2005**, *126*, 177–185. [[CrossRef](#)] [[PubMed](#)]
- Boggio, K.J.; Obasuyi, E.; Sugino, K.; Nelson, S.B.; Agar, N.Y.; Agar, J.N. Recent advances in single-cell MALDI mass spectrometry imaging and potential clinical impact. *Expert. Rev. Proteom.* **2011**, *8*, 591–604. [[CrossRef](#)] [[PubMed](#)]
- Kurczyk, A.; Gawin, M.; Chekan, M.; Wilk, A.; Łakomiec, K.; Mrukwa, G.; Fraczak, K.; Polanska, J.; Fajarewicz, K.; Pietrowska, M.; et al. Classification of thyroid tumors based on mass spectrometry imaging of tissue microarrays; a single-pixel approach. *Int. J. Mol. Sci.* **2020**, *21*, 6289. [[CrossRef](#)]
- Deininger, S.O.; Bollwein, C.; Casadonte, R.; Wandernoth, P.; Gonçalves, J.P.L.; Kriegsmann, K.; Kriegsmann, M.; Boskamp, T.; Kriegsmann, J.; Weichert, W.; et al. Multicenter Evaluation of Tissue Classification by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Imaging. *Anal. Chem.* **2022**, *94*, 8194–8201. [[CrossRef](#)]
- He, M.J.; Pu, W.; Wang, X.; Zhang, W.; Donge, T.; Dai, Y. Comparing DESI-MSI and MALDI-MSI mediated spatial metabolomics and their applications in cancer studies. *Front. Oncol.* **2022**, *12*, 891018. [[CrossRef](#)]
- Gonçalves, J.P.L.; Bollwein, C.; Schlitter, A.M.; Kriegsmann, M.; Jacob, A.; Weichert, W.; Schwamborn, K. MALDI-MSI: A Powerful Approach to Understand Primary Pancreatic Ductal Adenocarcinoma and Metastases. *Molecules* **2022**, *27*, 4811. [[CrossRef](#)]
- Capitoli, G.; Piga, I.; L'Imperio, V.; Clerici, F.; Leni, D.; Garancini, M.; Casati, G.; Galimberti, S.; Magni, F.; Pagni, F. Cytomolecular Classification of Thyroid Nodules Using Fine-Needle Washes Aspiration Biopsies. *Int. J. Mol. Sci.* **2022**, *23*, 4156. [[CrossRef](#)]
- Barajas-Solano, C.; Muñoz, B.; Chicano-Gálvez, E.; Escobar, P.; Mejía-Ospino, E. Discriminator for Cutaneous Leishmaniasis Using MALDI-MSI in a Murine Model. *J. Am. Soc. Mass Spectrom.* **2022**, *33*, 952–960. [[CrossRef](#)] [[PubMed](#)]
- Gibb, S.; Strimmer, K. MALDIquant: A versatile R package for the analysis of mass spectrometry data. *Bioinformatics* **2012**, *28*, 2270–2271. [[CrossRef](#)]
- Norris, J.L.; Cornett, D.S.; Mobley, J.A.; Andersson, M.; Seeley, E.H.; Chaurand, P.; Caprioli, R.M. Processing MALDI mass spectra to improve mass spectral direct tissue analysis. *Int. J. Mass Spectrom.* **2007**, *260*, 212–221. [[CrossRef](#)]
- Coomes, K.R.; Baggerly, K.A.; Morris, J.S. Pre-processing mass spectrometry data. In *Fundamentals of Data Mining in Genomics and Proteomics*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 79–102.
- Pelikan, R.C.; Hauskrecht, M. Automatic Selection of Preprocessing Methods for Improving Predictions on Mass Spectrometry Protein Profiles. In Proceedings of the AMIA Annual Symposium Proceedings, Washington, DC, USA, 13–17 November 2010; pp. 632–636.
- Pérez-Cova, M.; Bedia, C.; Stoll, D.R.; Tauler, R.; Jaumot, J. MSroi: A pre-processing tool for mass spectrometry-based studies. *Chemometr. Intell. Lab.* **2021**, *215*, 104333. [[CrossRef](#)]
- Ozcift, A.; Gulden, A. Assessing effects of pre-processing mass spectrometry data on classification performance. *Eur. J. Mass Spectrom.* **2008**, *14*, 267–273. [[CrossRef](#)] [[PubMed](#)]
- Cruz-Marcelo, A.; Guerra, R.; Vannucci, M.; Li, Y.; Lau, C.C.; Man, T.K. Comparison of algorithms for pre-processing of SELDI-TOF mass spectrometry data. *Bioinformatics* **2008**, *24*, 2129–2136. [[CrossRef](#)] [[PubMed](#)]
- Emanuele, V.A.; Gurbaxani, B.M. Benchmarking currently available SELDI-TOF MS preprocessing techniques. *Proteomics* **2009**, *9*, 1754–1762. [[CrossRef](#)] [[PubMed](#)]

17. Abdelmoula, W.M.; Lopez, B.G.C.; Randall, E.C.; Kapur, T.; Sarkaria, J.N.; White, F.M.; Agar, J.N.; Wells, W.M.; Agar, N.Y. Peak learning of mass spectrometry imaging data using artificial neural networks. *Nat. Commun.* **2021**, *12*, 5544. [[CrossRef](#)]
18. Lieb, F.; Boskamp, T.; Stark, H.G. Peak detection for MALDI mass spectrometry imaging data using sparse frame multipliers. *J. Proteom.* **2020**, *225*, 103852. [[CrossRef](#)]
19. Cleary, J.L.; Luu, G.T.; Pierce, E.C.; Dutton, R.J.; Sanchez, L.M. BLANKA: An Algorithm for blank subtraction in mass spectrometry of complex biological samples. *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 1426–1434. [[CrossRef](#)]
20. Wegdam, W.; Moerland, P.D.; Buist, M.R.; van Themaat, E.; Bleijlevens, B.; Hoefsloot, H.C.; de Koster, C.G.; Aerts, J.M. Classification-based comparison of pre-processing methods for interpretation of mass spectrometry generated clinical datasets. *Proteome Sci.* **2009**, *7*, 19. [[CrossRef](#)] [[PubMed](#)]
21. Oliveira, C.; Longuespée, R. MALDIImID: Spatialomics R package and Shiny app for more specific identification of MALDI imaging proteolytic peaks using LC-MS/MS-based proteomic biomarker discovery data. *Proteomics* **2023**, *23*, 2300005. [[CrossRef](#)]
22. Robichaud, G.; Garrard, K.P.; Barry, J.A.; Muddiman, D.C. MSiReader: An open-source interface to view and analyze high resolving power MS imaging files on Matlab platform. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 718–721. [[CrossRef](#)]
23. Jagadeesan, K.K.; Ekström, S. MALDIViz: A comprehensive informatics tool for MALDI-MS data visualization and analysis. *Slas Discov. Adv. Life Sci. R&D* **2017**, *22*, 1246–1252.
24. Bemis, K.A.; Föll, M.C.; Guo, D.; Lakkimsetty, S.S.; Vitek, O. Cardinal v. 3: A versatile open-source software for mass spectrometry imaging analysis. *Nat. Methods* **2023**, *20*, 1883–1886. [[CrossRef](#)] [[PubMed](#)]
25. Romano, P.; Profumo, A.; Facchiano, A. Pre-Processing MALDI/TOF Mass Spectra by Using Geena 2. *Curr. Protoc. Bioinform.* **2018**, *64*, e59. [[CrossRef](#)] [[PubMed](#)]
26. Capitoli, G.; Piga, I.; Clerici, F.; Brambilla, V.; Mahajneh, A.; Leni, D.; Garancini, M.; Pincelli, A.I.; L'Imperio, V.; Galimberti, S.; et al. Analysis of Hashimoto's thyroiditis on fine needle aspiration samples by MALDI-Imaging. *BBA-Proteins Proteom.* **2020**, *1868*, 140481. [[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.