PLOS ONE

RESEARCH ARTICLE

Multimodal MALDI imaging mass spectrometry for improved diagnosis of melanoma

Wanqiu Zhang[1,2]*, **Nathan Heath Patterson**[3,4], **Nico Verbeeck**[2], **Jessica L. Moore**[3¤a], **Alice Ly**[2], **Richard M. Caprioli**[3,4], **Bart De Moor**[1], **Jeremy L. Norris**[3,4¤b], **Marc Claesen**[2]

1 STADIUS Center for Dynamical Systems, Signal Processing, and Data Analytics, Department of Electrical Engineering (ESAT), KU Leuven, Leuven, Belgium, **2** Aspect Analytics NV, Genk, Belgium, **3** Frontier Diagnostics, LLC, Nashville, Tennessee, United States of America, **4** Mass Spectrometry Research Center, Department of Biochemistry, Vanderbilt University, Nashville, Tennessee, United States of America

¤a Current address: Discovery Life Sciences, Huntsville, Alabama, United States of America
¤b Current address: Bruker Scientific, Billerica, Massachusetts and Mass Spectrometry Research Center, Department of Biochemistry, Vanderbilt University, Nashville, Tennessee, United States of America
* wanqiu.zhang@aspect-analytics.com

Check for updates

OPEN ACCESS

Citation: Zhang W, Patterson NH, Verbeeck N, Moore JL, Ly A, Caprioli RM, et al. (2024) Multimodal MALDI imaging mass spectrometry for improved diagnosis of melanoma. PLoS ONE 19(5): e0304709. https://doi.org/10.1371/journal.pone.0304709

Editor: Francesco Savoia, IRCCS Istituto Romagnolo per lo Studio dei Tumori Dino Amadori / Skin Cancer Unit, Romagna Institute for the Study of Cancer IRST IRCCS Dino Amadori, ITALY

Received: September 20, 2023

Accepted: May 17, 2024

Published: May 31, 2024

Copyright: © 2024 Zhang et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data cannot be shared publicly because of patient data policy. Data are available from multiple institutions Duke University, and Indiana University for researchers who meet the criteria for access to confidential data. The data underlying the results presented in the study are available from Amanda Renick-Beech Administrative Officer, Mass Spectrometry Research Center Phone: 615-343-9207 Email: amanda.renick@Vanderbilt.Edu.

Abstract

Imaging mass spectrometry (IMS) provides promising avenues to augment histopathological investigation with rich spatio-molecular information. We have previously developed a classification model to differentiate melanoma from nevi lesions based on IMS protein data, a task that is challenging solely by histopathologic evaluation. Most IMS-focused studies collect microscopy in tandem with IMS data, but this microscopy data is generally omitted in downstream data analysis. Microscopy, nevertheless, forms the basis for traditional histopathology and thus contains invaluable morphological information. In this work, we developed a multimodal classification pipeline that uses deep learning, in the form of a pre-trained artificial neural network, to extract the meaningful morphological features from histopathological images, and combine it with the IMS data. To test whether this deep learning-based classification strategy can improve on our previous results in classification of melanocytic neoplasia, we utilized MALDI IMS data with collected serial H&E stained sections for 331 patients, and compared this multimodal classification pipeline to classifiers using either exclusively microscopy or IMS data. The multimodal pipeline achieved the best performance, with ROC-AUCs of 0.968 vs. 0.938 vs. 0.931 for the multimodal, unimodal microscopy and unimodal IMS pipelines respectively. Due to the use of a pre-trained network to perform the morphological feature extraction, this pipeline does not require any training on large amounts of microscopy data. As such, this framework can be readily applied to improve classification performance in other experimental settings where microscopy data is acquired in tandem with IMS experiments.

Introduction

Worldwide, skin cancer is the most common and the most deadly cancer in the white-skinned population [1, 2]. Currently, the diagnosis of melanoma is most frequently reliant upon

Funding: Wanqiu Zhang was supported by a Baekeland PhD grant from Flanders Innovation and Entrepreneurship (VLAIO) (https://www.vlaio.be/en/subsidies/baekeland-mandates). -This work was supported by Small Business Innovative Research Grant (https://www.sbir.gov/) No. R44CA228897-02, provided by the National Cancer Institute of the National Institutes of Health (https://www.nih.gov/). -KU Leuven (https://www.kuleuven.be/kuleuven): Research Fund (projects C16/15/059, C3/19/053, C24/18/022, C3/20/117, C3I-21-00316), Industrial Research Fund (Fellowships 13-0260,IOFm/16/004) and several Leuven Research and Development bilateral industrial projects. -Flemish Government Agencies: FWO (https://www.fwo.be/): EOS Project no G0F6718N (SeLMA), SBO project S005319N, Infrastructure project I013218N, TBM Project T001919N;PhD Grant (SB/1SA1319N), EWI (https://www.ewi-vlaanderen.be/en): the Flanders AI Research Program https://www.ewi-vlaanderen.be/en/flemish-ai-plan VLAIO (https://www.vlaio.be/en): CSBO (HBC.2021.0076) Baekeland PhD (HBC.20192204), - European Commission: European Research Council under the European Union's Horizon 2020 research and innovation programme (ERC Adv. Grant grant agreement No 885682) https://erc.europa.eu/apply-grant/starting-grant - Other funding: Foundation 'Kom op tegen Kanker', CM (Christelijke Mutualiteit)' https://www.cm.be/international.

Competing interests: JLM, NHP, SN, RMC, JLN, and JR disclose a financial interest in Fron- tier Diagnostics, LLC (FDx). FDx has issued and pending patent appli- cations in the US Patent Office that include part of the methods described in this paper. NV and MC, principals of Aspect Analytics NV, are paid consultants and provide services to FDx. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

histopathologic evaluation by dermatopathologists, which is inherently challenging. Several studies have reported significant discordance among experts in the diagnosis of challenging melanocytic lesions [3, 4]. For some challenging lesions, even experienced dermatopathologists had difficulty precisely or reproducibly classifying them into a certain category. Therefore, melanocytic lesions are the most common type of lesions that require a second opinion in pathology [5]. However, second opinions can be problematic, and often do not yield the needed accuracy.

One strategy to assist with melanoma diagnosis is by incorporating molecular technologies, to move to a more objective, evidence-based assessment. Imaging Mass Spectrometry (IMS) has been employed increasingly in clinical research and integrated into molecular pathology to complement routine histopathology evaluation [6]. The success of IMS is owed to its ability to characterize tissue microenvironments and provide a crucial link between tissue morphology and its molecular physiology. Many studies using IMS on the skin have already shown the potential in diagnosis of melanoma [7–15]. Many studies have employed IMS-technology to search for differences in protein profiles for melanoma diagnosis. Guran et al. [12] used an animal model to investigate the differences between melanoma and healthy skin protein profiles via MALDI IMS. Multivariate statistical analysis revealed m/z ions that are potentially involved in spontaneous regression of the disease. In order to identify proteomic differences between melanoma and benign melanocytic nevi, Casadonte et al. [7] applied IMS in a large cohort (n = 203) of cutaneous melanoma and nevi. The prediction model was tested in an independent set (n = 90) which achieved an accuracy of 93% in the classification of melanoma from nevi. Similarly, to discriminate between Spitz nevi and melanoma, Lazova et al. [16] performed IMS on 114 cases and correctly classified Spitz nevi with 97% sensitivity and 90% specificity in the independent sample set. Recently, Casadonte et al. [17] investigated and compared the molecular profile of mutated and wildtype melanoma to identify specific molecular signatures with the respective tumors via IMS-based proteomics technology. Results showed molecular differences between the two BRAF/NRAS mutated melanoma could be used to classify samples with an accuracy of 87%-89% and 76%-79% based on the classification models applied (i.e., linear discriminant analysis and support vector machine). A retrospective collaborative study (n = 102) reported by Lazova et al. [18] shows that IMS is more accurate than histopathology in the classification of diagnostically challenging, atypical Spitzoid neoplasms.

In a previous study [8], we utilized matrix-assisted laser desorption ionization imaging mass spectrometry (MALDI IMS) technology to classify melanoma and nevi lesions on a large cohort (n = 333), with the results showing high concordance (97.6% sensitivity and 96.4% specificity) with triconcordant histopathological evaluation by board-certified dermatopathologists. MALDI IMS is a beneficial alternative technology for skin cancer diagnosis as it uses tissue sections prepared in a similar manner as standard histopathology with minimal sample material for *in situ* analysis (a single 6 μm tissue section) and can a provide spatial information when using a serial section for histopathological annotation [8].

In our previous study, the pipeline we constructed solely used MALDI IMS data to predict the sample classification. However, for each of the IMS experiments, the pathologists used an H&E stain of the neighboring tissue section to guide the histopathological annotations. These H&E images provide crucial morphological information that pathologists use to diagnose melanoma lesions. By targeting the analytical measurement to regions with the highest probability of malignancy, we strengthen the classification algorithm. We hypothesize that a classification model that combines the H&E imaging data, which is the basis of histopathological diagnosis, with MALDI IMS data, an objective but highly accurate means of melanoma classification, that the overall classification accuracy could be further improved.

Multimodal imaging is a powerful tool which integrates information from different imaging techniques, and the resulting integrated datasets may open new opportunities for data mining. Some multimodal studies have shown some promising results in melanoma diagnosis. For example, for the detection of malignant melanoma, Wang et al. [19] proposed a multimodal framework to fuse clinical and dermoscopic images via a multi-label deep learning-based feature extractor and clinically constrained classifier chain. An average accuracy of 81.3% for detecting all labels and melanoma was achieved, which outperforms state-of-the-art methods by 6.4% or more. Recently, Schneider et al. [20] reported their multimodal approach for predicting BRAF mutation status in melanoma by combining individually trained random forest models on images, clinical and methylation data. This multimodal approach outperformed the unimodal models on an independent set: AUROC 0.80 (multimodal) vs. 0.63 (histopathologic image data) vs. 0.66 (clinical data) vs. 0.66 (methylation data). Although there is no reported study yet about IMS-based multimodal approach for melanoma diagnosis, this can be beneficial for MALDI IMS data, as they contain a substantial amount of chemical information but not enough readily available morphological knowledge from the tissue [21]. A multimodal MALDI IMS strategy often involves careful data acquisition, registration and computational data analysis to get the most/best of the different modalities [21]. In this way, complementarities and correlations from both modalities can be explored, providing better evidence towards biological understanding.

Currently, there are very few studies combining both spatial and chemical modalities, i.e., histopathological data and IMS data, to improve the downstream classification task (i.e., diagnosis). At present, there is one study that reports the synergy of metabolomics data from MALDI IMS and morphometry from histopathological data to improve the classification of kidney cancer [22]. This study was performed on a large patient cohort (n = 853) using FFPE tissue samples; when the classifier was trained on combined or multimodal datasets (morphometric and metabolite), it outperformed the classifiers trained on the individual data for each tumor subtype. This study shows the potential of using both histopathology and IMS data to improve the downstream diagnosis and tumor subtyping.

In comparison to the work by Prade et al. [22] which used a computer-assisted morphometry image analysis algorithm [23], we applied a pre-trained deep learning-based model to extract the morphological features from the histopathology dataset in this current study. Specifically, we adapted the pre-trained contrastive self-supervised learning model SimCLR [24], pre-trained on 57 histopathology datasets without any labels [25], to extract the high-level morphological features of the histopathological images previously only used to guide the MALDI IMS data acquisition [8].

In the machine learning field, the number of labeled data is found to have a positive correlation with task performance. However, the most tedious and time-consuming step is to annotate the histological data using medical image analysis applications. As labeled data is not often available, many approaches have been proposed to mitigate this bottleneck. For example, unsupervised learning [26] and its subclass self-supervised learning do not require annotations of the data, whereas semi-supervised learning uses partially labeled data. Interestingly, self-supervised approaches obtain 'labels' from the data itself and predict part of the data from other parts. Instead of focusing on finding specific patterns from the data like unsupervised learning often does, e.g., via clustering, self-supervised learning methods concentrate on recovering whole or part of some features from its original input, while still fitting in the type of supervised settings [27].

In this paper we adopted the pre-trained contrastive self-supervised learning model from Ciga et al. [25] and applied it on our previously collected 331 patients histopathology datasets [8]. Whereas in Al-Rohil et al. [8], the histopathological images were used to only guide MALDI-IMS data acquisition, here we also used them in the multimodal data analysis to improve

the downstream classification task. Our method successfully extracts relevant and highly qualitative morphological features from the histological images, without requiring any training nor labels. Finally, we combine the extracted morphological feature with MALDI IMS data in the downstream classification task to distinguish melanoma from nevus. We compare the classification performance of the unimodal and multimodal strategies. Our multimodal approach shows improvement over the unimodal MALDI IMS classification results, as reported in our previous work [8]. Our approach shows the potential of the synergism of spatial proteomics and morphology in (melanoma) cancer diagnosis.

Materials and methods

Tissue collection

The tissue samples used in this study were the same cohort as used in Al-Rohil et al. [8]. Deidentified FFPE skin biopsies from patients diagnosed with invasive melanoma and nevi were provided by the co-operating academic institutions and private practices.

All formalin fixed, paraffin embedded skin biopsies received from July 1, 2008 through July 1, 2019 were eligible for inclusion of the study. Individual specimens were selected by collaborating institutions. Study authors had no knowledge of patient identity and did not influence the selection of individual patients. Authors did not have access to information that could identify individual participants. Deidentified samples were provided by academic institutions and private practices. This includes Vanderbilt University, Vanderbilt University Medical Center, pathology Associates of St. Thomas, Duke University Medical Center, and Indiana University School of Medicine.

The review boards of Indiana University (IRB 1910325060), Vanderbilt University (IRB 030220), Duke University (IRB Pro00102363) and Sterling Institutional Review Board (Atlanta, GA) determined this study exempt from IRB review under the terms of the US Department of Health and Human Services Policy for Protection of Human Research Subjects at 45 CFR §46.104(d).

Problem statement

The diagnosis of melanoma relies heavily on histopathology evaluation. However, significant discordance among the experts occurs when it comes to challenging melanocytic lesions. Utilizing molecular technologies such as Imaging Mass Spectrometry (IMS) can facilitate a more objective, evidence-based assessment.

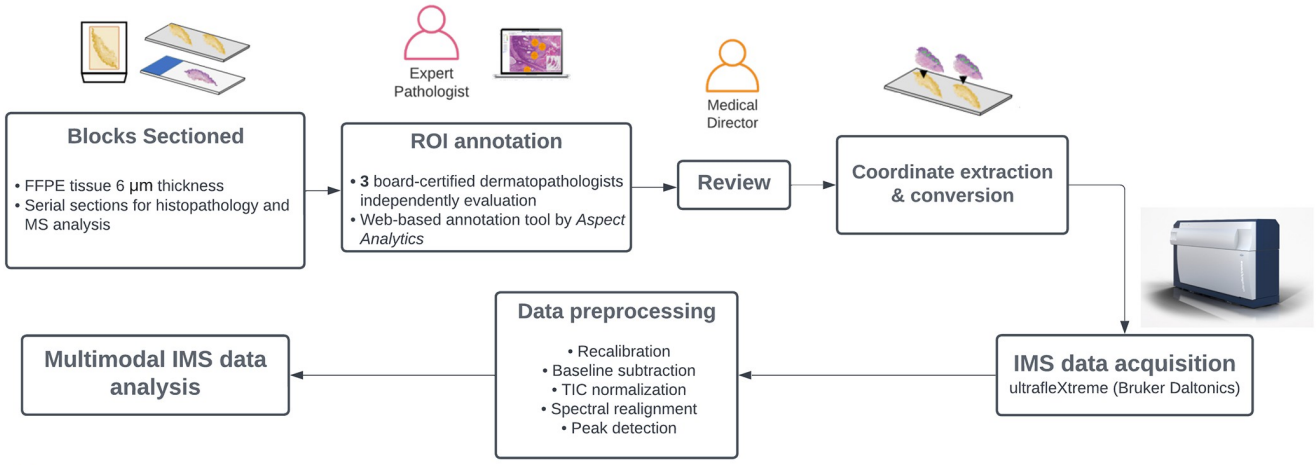
Presently, many IMS-based studies do not fully exploit the accompanying microscopy data, which is almost invariably collected with IMS data, in their downstream analyses. Microscopy, as the basis of histopathology evaluation, provides invaluable morphological information, however this is typically ignored in MSI studies, with the microscopy data typically only used to select regions of interest e.g. for building classifiers.

We propose a multimodal pipeline that incorporates IMS data and microscopy, which improves the downstream data analysis results, particularly in diagnosis and classification. This approach is especially useful for histopathologically challenging cases where IMS technology is performed.

Histopathology sample preparation and data acquisition

Serial sections (6 μm) were cut from deidentified patient biopsy FFPE tissue blocks and used for histopathology and mass spectrometry analysis. Fig 1(a) illustrates the workflow of histology-guided IMS sample preparation and processing.

(a)



(b)

Final multimodal IMS and microscopy data		
	Training set	**Held-out test set**
Patients	239	92
Spots	5080	1924
Diagnosis Melanoma spots	2476	999
Diagnosis Nevus spots	2604	925

Fig 1. Overview of the histology-guided IMS sample preparation and processing workflow (a) and information about the final included number of samples and spots (b). The workflow (a) leads into multimodal data analysis and classification, incorporating both chemical information from MALDI IMS and morphological information from H&E microscopy.

https://doi.org/10.1371/journal.pone.0304709.g001

Sections used for histopathological analysis were prepared and H&E stained according to standardized institutional practices and scanned at 20x magnification using a Leica SCN-400 digital slide scanner (Leica Biosystems, Buffalo Grove, IL, USA). The digitized sections were independently annotated by three board-certified dermatopathologists in a blinded fashion using web-based annotation tools: PIMS [28] and Annotation studio (Aspect Analytics NV, Genk, Belgium). Each sample was provided to the pathologist with the patient demographic and with the goal of classifying the tissue as either melanoma or nevus.

Using the annotation software, the dermatopathologists annotated regions of interest from each sample/slide with a 50 µm diameter. While 6 different annotation labels were placed (melanoma *in situ*, invasive melanoma, junctional nevus, intradermal nevus, uninvolved epidermis and dermal stroma), this study focused only on building a binary classifier to distinguish the melanoma class from nevus class.

MALDI IMS sample preparation and data acquisition

The sample preparation is the same as that described by Al-Rohil et al. [8]. All sections for MALDI IMS were mounted on ITO slides (Delta Technologies, Loveland, CO). Briefly, sections were incubated in an oven at 55˚C for 1 hour, deparaffinized (xylene x 2), rehydrated in ethanol (100% x 2, 95%, 70%), washed in ultrapure water (x2) and dried at 37˚C. Antigen retrieval was conducted in EDTA buffer (pH 8.5) in an Instant Pot Electric Pressure cooker (IP-Duo60 V3) for 10 min on high pressure mode. For *in situ* tryptic digestion, sections were

coated with 0.074 μg/mL mass spectrometry-grade trypsin (Sigma Aldrich) in a 100 mM ammonium bicarbonate buffer solution using a M5-Sprayer (HTX Technologies, Chapel Hill, NC), and incubated at 37˚C overnight. After digestion, the samples were coated in 5 mg/mL α-cyano-4-hydroxycinnamic acid matrix in 90% acetonitrile/deionized water and 0.2% TFA using the M5-Sprayer.

The dermatopathologist annotations were used to generate coordinates to guide the MALDI IMS acquisition from the same region/s on the unstained serial sections. MALDI IMS was performed on an ultrafleXtreme MALDI-TOF MS (Bruker Daltonics, Billerica, MA, USA) fitted with a SmartBeam Laser Nd:YAG laser following previously published parameters [8]. Data was collected in positive reflector mode over the range of 700 to 3500 Daltons, at 500 laser shots per each profile spectrum for each region of interest.

Data processing and sample selection

A series of preprocessing methods were applied on the acquired IMS data, including resampling of the data to a common m/z axis, baseline subtraction, TIC normalization, spectral realignment and peak detection [8].

For the multi-modal image analysis, we only included tri-concordant samples (n = 331), meaning all dermatopathologists agreed with the diagnosis. Samples that were either diagnostically challenging, or did not meet the minimal mass spectral quality standards were excluded. Cases where only one modality is available (n = 2) were also excluded from this study. Other than that, we used the same samples as the previously reported study by Al-Rohil et al. [8], and can compare the multimodal results with the unimodal IMS classification reported therein.

The included samples contained 21 collected spots on average. As we focused on the binary classification task between melanoma and nevus, spots with annotations as melanoma (i.e., invasive melanoma and melanoma *in situ*) and nevus (i.e., junctional nevus and intradermal nevus) were included (n = 7004). It must be noted that, in this study, we measured the classification performance only at the spots level and not at the aggregated patients/samples level. Our spot-level results can be directly compared to those reported in Al-Rohil et al. [8].

The final multimodal IMS and microscopy dataset was randomly split into a training and a held-out test set at the patient level in order to prevent information leaks. Specifically, we designed our splits such that training and held-out test data spots originate from different patients, to ensure we are accurately assessing generalization performance on unseen slides, both during cross-validation fold generation (in the training data) as well as in determining the training vs held-out sets. The held-out test set in this study means that the data was never used during the whole model training part (during the parameter tuning, model selection). Fig 1(b) shows the information of the final included number of patients and the collected spots.

Data analysis

In this study, we built 3 different classification pipelines for melanoma diagnosis, namely the *unimodal MALDI IMS*, the *unimodal microscopy*, and the *multimodal pipeline*. As illustrated in Fig 2, the downstream classification part is the same across all compared pipelines but the inputs are different. For unimodal pipelines, only a single modality, either IMS data or microscopy, is utilized as the input for the downstream classification task, whereas in the multimodal setting, both modalities are used as the input for classification model training. Finally, these 3 pipelines are compared based on their classification performance.

Unimodal IMS. Prior to feeding to the classification model, IMS data was preprocessed and peak picked. Each mass spectrum was generated from a single spot, which has 50 μm. After peak picking, a total number of 5558 peaks remained. Then we feed the preprocessed

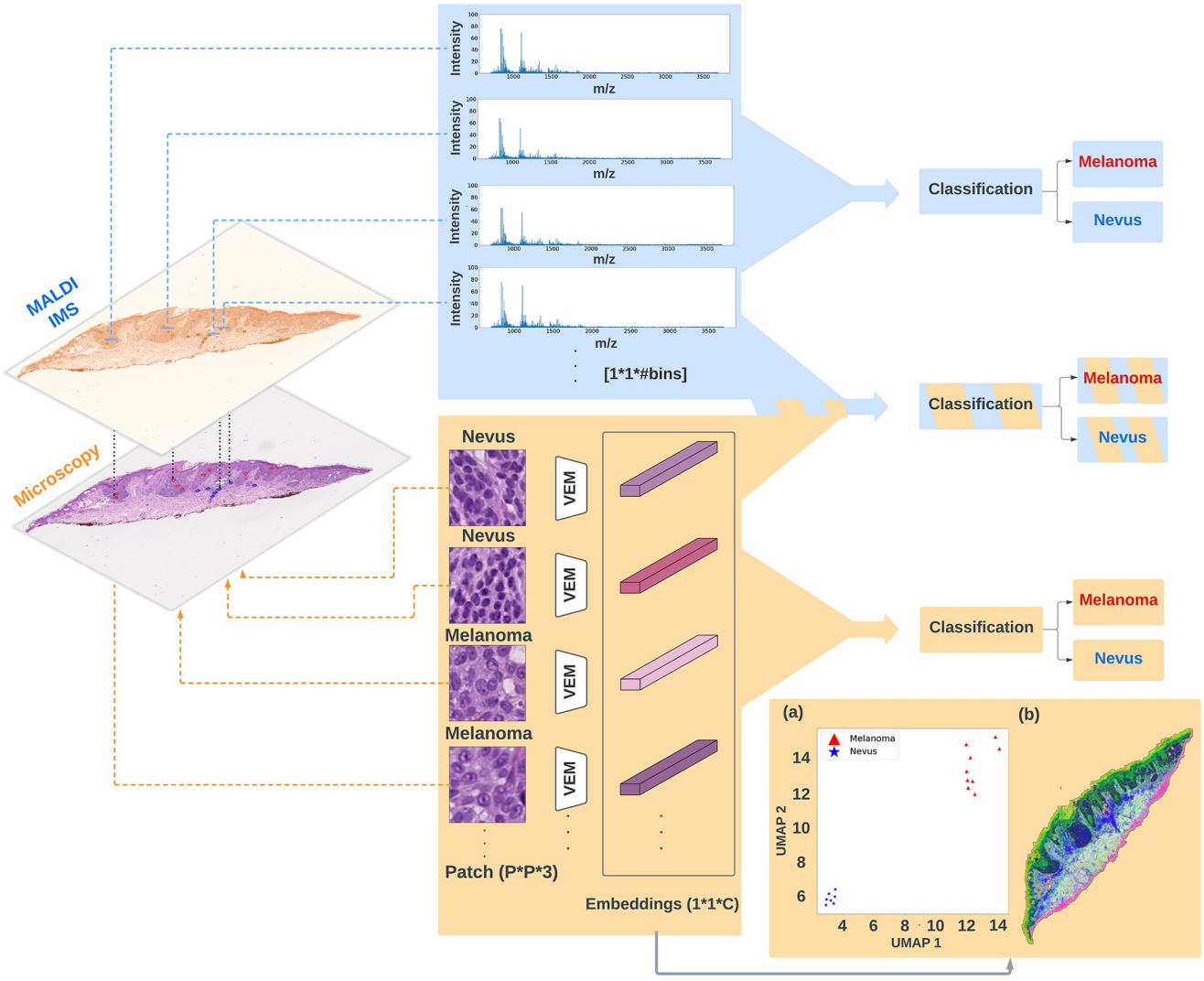


Fig 2. Data analysis workflow on an example tissue: **top blue panel** represents the unimodal IMS classification pipeline, where only IMS data are used for the downstream classification task; **bottom orange panel** shows the unimodal microscopy classification pipeline, where only morphology features are included for the melanoma diagnosis; In addition, intermediate embeddings from unimodal microscopy model were visualized via UMAP method: (a) 2-D embeddings of 16 patches (measured spots) from this example tissue with assigned colors based on pathologists' annotations; (b) hyperspectral visualization of 3-D embeddings from all patches across the whole example tissue; **middle fused panel** is the multimodal strategy, where both IMS data and morphology features are used to distinguish melanoma from nevus *#bins = 5558, P = 96, C = 512, VEM: vector embedding morphology model.*

https://doi.org/10.1371/journal.pone.0304709.g002

IMS data to the downstream classification model. This pipeline was reproduced from our previous study [8] and compared with the multimodal pipeline in this study.

Unimodal microscopy. We applied a pre-trained deep learning-based, self-supervised model to extract the features from microscopy data [25]. Each 3-color-channel histopathological image is transformed to a single dense vector and then fed to the downstream classification task. This pre-trained model was adapted originally from Chen et al. [24], where the model learns the feature from data by maximizing the agreement between two stochastically augmented views of the same image via a contrastive loss function. Augmentation plays an essential role in contrastive learning. In the context of the histopathology dataset, some augmentation algorithms from the original paper [24] may not be suitable here, as it was

applied on natural-scene images like ImageNet [29]. Thus, Ciga et al. [25] conducted a series of experiments and some minor modifications on data augmentation. In the end, the augmentation algorithms included in the study are listed as following: randomly resized crops, 90 degrees rotations, horizontal and vertical flips, color jittering and Gaussian blurring.

Ciga et al. [25] pre-trained their model on 57 multiple multi-organ histopathology datasets without any labels, as it is a self-supervised learning model. Those datasets contain various types of staining and resolutions, which helps the model learn better quality features. As reported in Ciga et al. [25], the features learned from the model that was pre-trained on those histopathology datasets, are better than the features that were learned from networks pre-trained on ImageNet. In total, more than 4 millions of patches (in pixel size of 224*224) are extracted from histopathology datasets and then fed to the model training.

In this study, we adapted this self-supervised model, pre-trained on a large number of histopathologic images, to the previously collected H&E stained histopathological images from our previous study [8]. Each spot generates a single unique mass spectrum, and has a diameter of 50 μm. Thus, each patch in the histopathology dataset has a physical size of 48 μm*48 μm. The collection of histopathological patches consists of mainly of 20x resolution samples (96*96 pixel size), and a small portion of 40x resolution samples (192*192 pixel size). We applied some preprocessing steps on images following the suggestions in [25], including resizing, and normalization. Finally, each microscopy patch in the dataset is embedded into a vector representation via this pre-trained deep learning based self-supervised learning model, referred to as a *'Vector Embedding Morphology'* (VEM) model in this paper. The resulting embeddings contain 512 dimensions. In order to visualize high-dimensional abstract features, we applied the method Uniform Manifold Approximation (UMAP) [30] to project the data into lower dimensions. UMAP is a dimensionality reduction method, which seeks to search for a low-dimensional representation of the data that has the closest possible fuzzy topological representation of the original high-dimensional data. Previously, Smets et al. [31] evaluated the performance of the UMAP method on IMS data and compared it with other popular dimensionality reduction algorithms like t-distributed stochastic neighbor embedding (t-SNE) and principal component analysis (PCA).

Multimodal IMS and microscopy. In the multimodal setting, we represent the data for each spot with a single vector, which is created by concatenating the preprocessed IMS data for a spot together with the extracted morphological features from microscopy data for that same spot, with equal weights. In this way we achieve a fusing of the two modalities, which serves as an input for the classification algorithm.

Classification model. We applied Support Vector Machines (SVMs) [32, 33] to build the binary classifiers for each pipeline. As a popular supervised learning model, SVMs try to find a hyperplane as a decision boundary to separate data points with different classes, where this hyperplane has the maximal distance between data with different labels. To implement the SVMs algorithm, we used the Python package Scikit-learn [34] with a linear kernel.

Model selection and evaluation. During the parameter tuning for SVMs models, we used a grid search strategy, where the model exhaustively generates candidates from a grid of parameter values that are predefined by the user S1 File. Furthermore, we leveraged *nested cross-validation* in order to estimate the generalization error underlying each model including the performance variability induced by the hyperparameter optimization process itself. Additionally, in order to further prevent information leaking, spots measured from the same patient/tissue were grouped in the same fold during the whole nest cross-validation process. In summary, we first split the total samples into a training set and a held-out test set on a patient level. Then we applied a grouped-nested-cross-validation strategy on the training set for model selection and evaluation. The held-out test set was not used during the whole training

procedure. Thus the whole framework was designed carefully during the model selection and evaluation, so that we can fairly compare classification performance from unimodal and multimodal pipelines.

Results

In this study, we built 3 different pipelines namely the unimodal microscopy, the unimodal IMS and the multimodal classification pipelines. Unimodal pipelines only use one modality i.e., either IMS or microscopy data for the downstream melanoma diagnosis, while multimodal strategy utilized both modalities. In order to compare those 3 pipelines, we conducted various measurements to assess their classification performance. In addition, we also applied a dimensionality reduction method UMAP to visualize the intermediate data in lower dimensions.

UMAP visualization of intermediate data

In order to provide more intuition for these highly abstract features, we used UMAP to project and visualize the extracted morphology feature from the unimodal microscopy pipeline. As shown in Fig 2(a), after the feature extraction, each of those 16 patches (measured spots from the example tissue) is transformed into a single vector and shown in a 2-d scatter plot. Each dot in the scatter plot represents a vector representation of a microscopy patch and is assigned a red or blue color based on the pathology annotations either melanoma or nevus, respectively. Clearly, all melanoma patches are grouped together and far away from the cluster of nevus patches. Additionally, we collected all patches across the whole slide of this example tissue and used UMAP to project their 512-d transformed morphological embeddings into 3-d feature space. The resulting 3-d UMAP embeddings are visualized via translating each location of the patch to the RGB color scheme, which is referred to as hyperspectral visualization and shown in Fig 2(b). Regions where share similar colors represent their morphology are similar with each other. These results show the potential of the morphology features learned from *'Vector Embedding Morphology'* (VEM) used in the unimodal microscopy pipeline, and provide more interpretation of these abstract embeddings. Importantly, this feature extraction process requires no training nor labels.

Furthermore, we performed UMAP on each input data from unimodal and multimodal pipelines. Three different inputs are collected and compared in this study, namely, (1) preprocessed IMS data (n_dimension = 5558), (2) morphological features extracted from microscopy data via a pre-trained self-supervised learning model (n_dimension = 512), (3) the combination of normalized (1) and (2) with equal weight (n_dimension = 5558+512 = 6070). UMAP was conducted to map those high-dimensional inputs to 2 dimensions in an unsupervised manner. Additionally, we also provide the labels for each data point in order to have a better intuition of the representations from those different input data.

As shown in Fig 3, each dot in the scatter plot represents a mass spectrum or a vector representation of a histopathological image or the combination of both, with its color showing the annotation provided by 3 dermatopathologists. Each column shows the UMAP embeddings of features from microscopy data, IMS and the combined data, respectively from left to right. The first and second rows represent the input data from the training dataset and the independent held-out test set respectively. We applied UMAP via Python implementation [35] with cosine distance as metric for each type of input data.

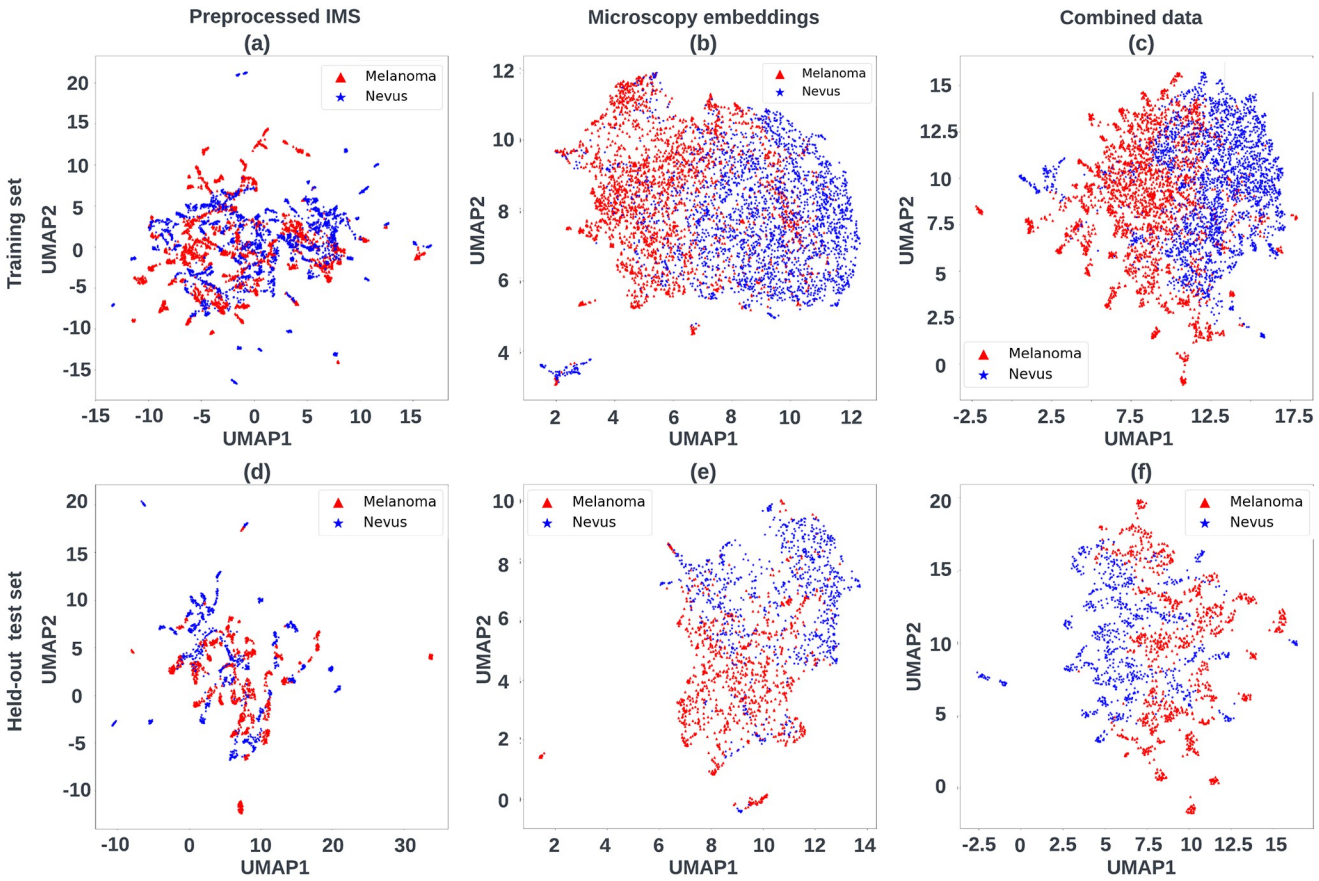


Fig 3. UMAP 2-D visualization of each input data across pipelines (unsupervised). In the first row (a) preprocessed IMS data, (b) extracted morphological features from microscopy data and (c) an equally weighted combination of (a) and (b) are from the training dataset. The second row shows the UMAP visualization from the held-out test dataset. The assigned color of each spot is decided by its diagnosis from 3 dermatopathologists.

https://doi.org/10.1371/journal.pone.0304709.g003

Classification performance

In this study we applied various classification performance measurements such as, Receiver operating characteristic—Area under the curve (ROC-AUC), F1, precision and recall scores. The range of those metrics is typically from 0 to 1, where 1 represents the perfect classification performance. ROC curves are widely used to measure performance in binary classification tasks, as it plots the true positive rates (TPR) and the false positive rate (FPR) with different thresholds. By computing the area under the ROC curve, the performance across the full range of classification thresholds can be summarized in a single number.

As mentioned earlier, we carefully performed the *grouped-nest-cross-validation* strategy throughout the modeling process in order to avoid any possible information leaking. We applied this strategy on the training dataset and fit the optimized classification model for each pipeline on an independent held-out test set. The results are shown in Fig 4.

We reproduced the unimodal MALDI IMS classification results from our previous study [8], and obtained comparable ROC-AUC results (0.910 vs. 0.915 on training set, 0.933 vs. 0.938 on the held-out test set, for reported spots-level results from Al-Rohil et al. [8] and our reproduced results, respectively). Interestingly, the unimodal microscopy pipeline achieves higher mean ROC-AUC scores than the unimodal IMS (0.937 vs. 0.915) on the training

(a) Classification performance from training and held-out test dataset

Model	ROC-AUC		F1 score		Precision		Recall	
	Training set	Held-out test set	Training set	Held-out test set	Training set	Held-out test set	Training set	Held-out test set
Unimodal IMS	0.915 (0.049)	0.931	0.823 (0.067)	0.856	0.846 (0.085)	0.828	0.815 (0.11)	0.886
Unimodal microscopy	0.937 (0.03)	0.938	0.82 (0.056)	0.861	0.857 (0.104)	0.871	0.805 (0.111)	0.851
Multimodal	**0.968** (0.023)	**0.968**	**0.866** (0.056)	**0.91**	**0.920** (0.051)	**0.924**	**0.83** (0.112)	**0.90**

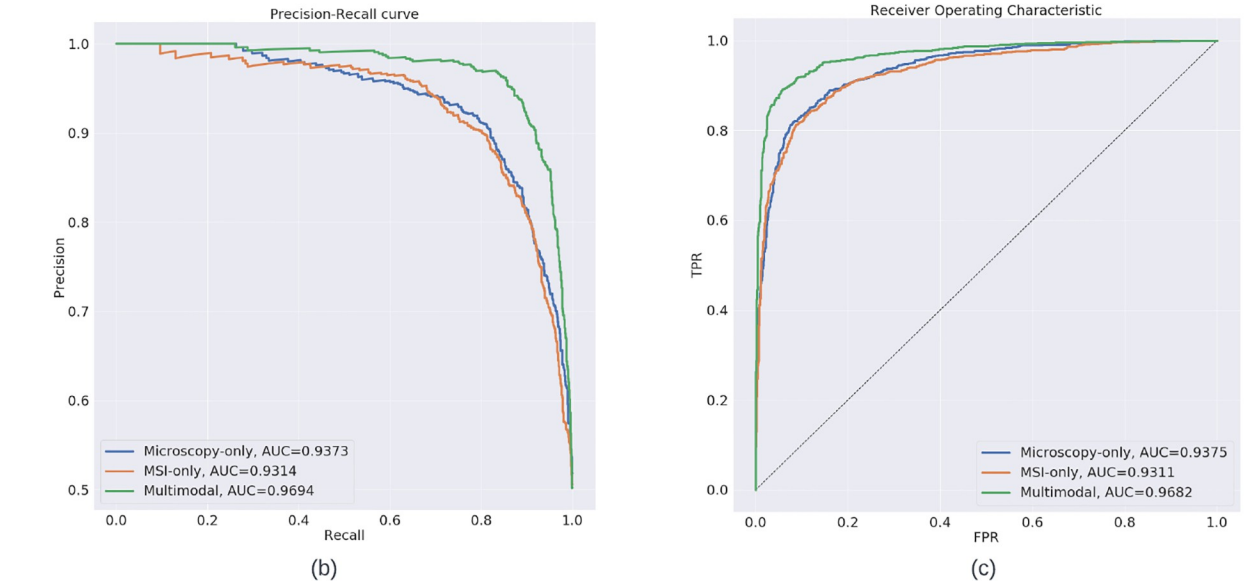


Fig 4. Classification results on training and held-out test sets. All results are based on **spot-level**. (a) shows the classification performance of each pipeline on training and held-out test dataset. The results from the training set represent the mean value of the performance (with their standard deviation) after the nested-cross-validation process. (b)and(c) show the precision-recall and ROC cures plots of each pipeline on the **held-out test set**.

https://doi.org/10.1371/journal.pone.0304709.g004

dataset, while achieving comparable ROC-AUC results on the held-out test set. The result shows the advantage of using a pre-trained deep learning model for extracting important morphological features from microscopy data. However, it must be noted that the annotations used to train and evaluate the classification model are fully based on microscopy data by dermatopathologists.

By combining information originating from the two available modalities in a single classifier, i.e. the multimodal pipeline, we obtain the best classification performance, (ROC-AUC as 0.968 on training and held-out test set). Additionally, in order to check whether the improvement presented by our proposed multimodal pipeline was significant compared to the previous study [8] (i.e., unimodal IMS pipeline), we applied the DeLong test to compare two AUC results obtained from the unimodal IMS and multimodal pipelines on the independent held-out test set. The resulting p-value (2.2e-16) confirms the significance of the improvement using the multimodal pipeline. This demonstrates the synergy in combining observations in the biochemical domain, obtained through IMS, with those in the morphology domain, which have been traditionally used in melanoma diagnosis.

The unimodal IMS and unimodal microscopy pipelines show comparable F1 scores across both training and held-out test sets. Unimodal IMS outperforms unimodal microscopy on

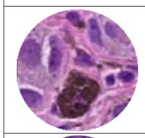
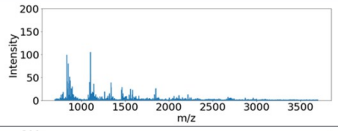
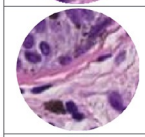
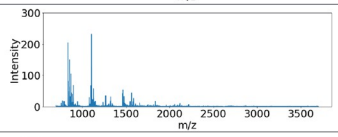
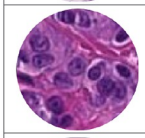
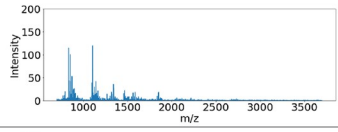
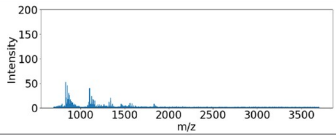
Microscopy data	Imaging mass spectrometry data	Unimodal microscopy prediction	Unimodal IMS prediction	Multimodal prediction	Experts' evaluation
		Nevus	Melanoma	Melanoma	Melanoma
		Nevus	Melanoma	Nevus	Nevus
		Melanoma	Melanoma	Nevus	Nevus
		Nevus	Nevus	Melanoma	Melanoma

Fig 5. Misclassified cases by unimodal algorithms from the held-out test set. Red and green text indicate model results that are incongruent and congruent with expert pathology assessment, respectively.

https://doi.org/10.1371/journal.pone.0304709.g005

both training and held-out test sets based on the recall scores (0.815 vs. 0.805, 0.886 vs. 0.851). On the other hand, unimodal microscopy achieves higher precision scores than unimodal IMS on training and held-out test sets. These slight variations in terms of which approach is best are attributable to differences in classification thresholds, thus emphasizing the importance of using an overall metric like ROC-AUC that accounts for the classifiers' full range.

We also computed the standard deviation for each score based on the training data, and the multimodal pipeline achieves the lowest values almost across all metrics. For example, standard deviations from ROC-AUC: 0.023 (multimodal) vs 0.03 (unimodal microscopy) vs 0.049 (unimodal IMS). This indicates higher stability of the multimodal model, which additionally reinforces the value of combining both modalities.

Additionally, we plotted the Precision-Recall (PR) curve and ROC curves from the independent held-out test set in Fig 4(b) and 4(c) respectively. PR curves illustrate the trade-off between precision score and recall score with different thresholds. Similarly to ROC-AUC, higher area under the PR curve (PR-AUC) represents higher recall and precision across all classification thresholds. PR and ROC curves are typically used to assess the binary classification performance. As shown in Fig 4(b) and 4(c), the multimodal pipeline achieves the highest PR-AUC and ROC-AUC scores (0.969, 0.968).

Furthermore, we encountered some interesting misclassified cases on the held-out test dataset. As shown in Fig 5, the multimodal strategy shows its power in accurate classification of melanoma and nevus, even when both unimodal strategies fail to yield correct predictions. This highlights that our multimodal classification model is greater than the sum of its (unimodal) parts.

Discussion

This study extends and improves the previous melanoma diagnosis study by Al-Rohil et al. [8] by incorporating the morphological features from microscopy data in the classification model

along with MALDI IMS data. The spot-level ROC-AUC results from the classification model have been largely improved from previous study strategy namely unimodal MALDI IMS (0.915 vs. 0.968, training data), as well as the other scores such as F1 score, precision and recall. Additionally, the classification model training became more stable and robust with the help of multimodal strategy, evidenced by lower variability in model performance.

At present, most IMS classification studies largely focus on only using histopathology data to make annotations that guide the IMS data collection or restrict which spectra are used in order to build more accurate classifiers [36, 37]. Secondly, multimodal MALDI IMS analysis requires careful data acquisition, dedicated registration before the downstream data analysis, which is not always trivial. Last but not the least, learning high-quality morphological features from microscopy data could require a large amount of manpower for dedicated annotations and also introduce computational challenges.

To our knowledge, there is only one study [22] utilizing the information from histopathology data, via computer-aided morphometric-based image analysis, in combination with MALDI IMS to improve renal carcinoma subtyping. In comparison with their work, instead of using morphometric-based algorithms, we applied a deep learning-based self-surprised model (pre-trained on a large histopathology dataset) to extract the high-quality morphological features from microscopy data in a fully unsupervised manner. This model is highly efficient as there is no training required during the feature extraction. However, the extracted features are highly abstract, which is a common challenge in the deep learning field. Thus the resulting features from microscopy data are not as interpretable as the ones learned from the morphometric-based method as used by Prade et al. [22]. To improve the interpretability of the learned features, we applied the dimensionality reduction method like UMAP on those high-dimensional features (512) and projected them into 2 & 3-d feature space.

In the context of clinical practice, which is primarily evidence-based, transparency is required in decision making [38]. Lack of explainability is one of the key limitations for current many advanced AI algorithms, especially 'black-box' models. However, this is not only a technological issue, but also introduces a series of medical, legal, ethical and societal questions and discussions [39]. Recently, large efforts have been invested in explainable AI. In the field of computer vision, methods like Grad-CAM [40] can provide 'visual explanations' for decisions from 'black-box' models, i.e., neural networks. The future work for our study could be applying those explainable AI algorithms to the feature extraction process on microscopy data, making those extracted abstract features more interpretable. Moreover, despite the abstract nature of deep learning-based features, similar approaches have already been approved by the FDA for in vitro diagnostic (IVD) use [41]. On top of that, an increasing volume of comparable methods are being used for clinical decision support. As such, while the abstract nature of these features can be considered a downside, AI-based methods are already actively complementing (in case of decision support) or even surpass interpretable approaches (as per the example of Paige AI's IVD approved pipeline).

Finally, we built a linear classifier on the top of IMS features and/or morphological features to distinguish melanoma from nevus. The classification results show that a multimodal strategy, with a relatively simple equal weighted combination method, can outperform other unimodal workflows in both prediction accuracy and robustness. Microscopy images are acquired in tandem in a majority of IMS experiments, and as such this methodology provides excellent opportunities to improve classification and downstream machine learning at very little additional expense. Due to the use of pre-trained networks, which do not require additional training, this additional feature extraction step from microscopy images is near 'plug-and-play', and could be readily incorporated with IMS experiments on other types of tissues in the future.

Conclusion

With the increased prevalence of multimodal studies, especially those integrating traditional histological or cellular stains with molecular analyses, it is important to develop computational approaches that effectively combine the data to go beyond qualitative comparisons. A critical area of such combinations is clinical assays where the increase in information from modality combinations has the potential to improve health outcomes. In this work, we demonstrate the potential of a multimodal strategy on a relatively large patient cohort (n = 331) which combines morphological features extracted from standard clinical microscopy data and MALDI IMS molecular data the melanoma diagnosis. In future work, this strategy can be further validated on other types of cancer tissues, more advanced fusion strategies could be developed, and the approach could be applied to other combinations of morphological optical images and molecular data.

Supporting information

S1 File. Details of hyperparamter finetuning. Results of the nested cross-validation on the training dataset and the fine tuned hypermeters.
(PDF)

Acknowledgments

The authors thank Ahmed Alomari, Rami Al-Rohil, Jason Robbins for for providing pathological analysis, and Thomas Moerman for software development.

Author Contributions

Conceptualization: Wanqiu Zhang, Nico Verbeeck, Marc Claesen.

Data curation: Wanqiu Zhang, Nathan Heath Patterson, Nico Verbeeck, Jessica L. Moore, Marc Claesen.

Formal analysis: Wanqiu Zhang, Nathan Heath Patterson, Nico Verbeeck, Marc Claesen.

Funding acquisition: Bart De Moor, Jeremy L. Norris, Marc Claesen.

Methodology: Wanqiu Zhang, Nathan Heath Patterson, Nico Verbeeck, Jessica L. Moore, Marc Claesen.

Software: Wanqiu Zhang, Nathan Heath Patterson, Nico Verbeeck, Marc Claesen.

Supervision: Richard M. Caprioli, Bart De Moor, Jeremy L. Norris.

Validation: Nico Verbeeck, Marc Claesen.

Visualization: Wanqiu Zhang, Nathan Heath Patterson.

Writing – original draft: Wanqiu Zhang, Nico Verbeeck, Alice Ly, Marc Claesen.

Writing – review & editing: Nathan Heath Patterson, Jessica L. Moore, Richard M. Caprioli, Bart De Moor, Jeremy L. Norris.

References

1. Available from: https://www.cancer.org/cancer/melanoma-skin-cancer/about/key-statistics.html. Date of access: 2022-11-24.
2. Leiter U, Keim U, Garbe C. Epidemiology of skin cancer: update 2019. Sunlight, Vitamin D and Skin Cancer. 2020; p. 123–139. https://doi.org/10.1007/978-3-030-46227-7_6 PMID: 32918216

3. Gerami P, Jewell SS, Morrison LE, Blondin B, Schulz J, Ruffalo T, et al. Fluorescence in situ hybridization (FISH) as an ancillary diagnostic tool in the diagnosis of melanoma. Am J Surg Pathol. 2009; 33 (8):1146–1156. https://doi.org/10.1097/PAS.0b013e3181a1ef36 PMID: 19561450

4. Veenhuizen KC, De Wit PE, Mooi WJ, Scheffer E, Verbeek AL, Ruiter DJ. Quality assessment by expert opinion in melanoma pathology: experience of the pathology panel of the Dutch Melanoma Working Party. J Pathol. 1997; 182(3):266–272. https://doi.org/10.1002/(SICI)1096-9896(199707)182:3%3C266::AID-PATH812%3E3.0.CO;2-%23 PMID: 9349228

5. Zembowicz A, Scolyer RA. Nevus/melanocytoma/melanoma: an emerging paradigm for classification of melanocytic neoplasms? Arch Pathol Lab Med. 2011; 135(3):300–306. PMID: 21366452

6. Porta Siegel T, Hamm G, Bunch J, Cappell J, Fletcher JS, Schwamborn K. Mass spectrometry imaging and integration with other imaging modalities for greater molecular understanding of biological tissues. Mol Imaging Biol. 2018; 20(6):888–901. https://doi.org/10.1007/s11307-018-1267-y PMID: 30167993

7. Casadonte R, Kriegsmann M, Kriegsmann K, Hauk I, Meliss RR, Mueller CS, et al. Imaging mass spectrometry-based proteomic analysis to differentiate melanocytic nevi and malignant melanoma. Cancers. 2021; 13(13):3197. https://doi.org/10.3390/cancers13133197 PMID: 34206844

8. Al-Rohil RN, Moore JL, Patterson NH, Nicholson S, Verbeeck N, Claesen M, et al. Diagnosis of melanoma by imaging mass spectrometry: Development and validation of a melanoma prediction model. J Cutan Pathol. 2021; 48(12):1455–1462. https://doi.org/10.1111/cup.14083 PMID: 34151458

9. Lazova R, Seeley EH. Proteomic mass spectrometry imaging for skin cancer diagnosis. Dermatol Clin. 2017; 35(4):513–519. https://doi.org/10.1016/j.det.2017.06.012 PMID: 28886807

10. Taverna D, Nanney LB, Pollins AC, Sindona G, Caprioli R. Spatial mapping by imaging mass spectrometry offers advancements for rapid definition of human skin proteomic signatures. Exp Dermatol. 2011; 20(8):642–647. https://doi.org/10.1111/j.1600-0625.2011.01289.x PMID: 21545539

11. de Macedo CS, Anderson DM, Schey KL. MALDI (matrix assisted laser desorption ionization) imaging mass spectrometry (IMS) of skin: aspects of sample preparation. Talanta. 2017; 174:325–335. https://doi.org/10.1016/j.talanta.2017.06.018 PMID: 28738588

12. Guran R, Vanickova L, Horak V, Krizkova S, Michalek P, Heger Z, et al. MALDI MSI of MeLiM melanoma: Searching for differences in protein profiles. PLoS One. 2017; 12(12):e0189305. https://doi.org/10.1371/journal.pone.0189305 PMID: 29220390

13. Lazova R, Smoot K, Anderson H, Powell MJ, Rosenberg AS, Rongioletti F, et al. Histopathology-guided mass spectrometry differentiates benign nevi from malignant melanoma. J Cutan Pathol. 2020; 47 (3):226–240. https://doi.org/10.1111/cup.13610 PMID: 31697431

14. Alomari AK, Glusac EJ, Choi J, Hui P, Seeley EH, Caprioli RM, et al. Congenital nevi versus metastatic melanoma in a newborn to a mother with malignant melanoma–diagnosis supported by sex chromosome analysis and imaging mass spectrometry. J Cutan Pathol. 2015; 42(10):757–764. https://doi.org/10.1111/cup.12523 PMID: 25989266

15. Taverna D, Boraldi F, De Santis G, Caprioli RM, Quaglino D. Histology-directed and imaging mass spectrometry: An emerging technology in ectopic calcification. Bone. 2015; 74:83–94. https://doi.org/10.1016/j.bone.2015.01.004 PMID: 25595835

16. Lazova R, Seeley EH, Keenan M, Gueorguieva R, Caprioli RM. Imaging mass spectrometry–a new and promising method to differentiate Spitz nevi from Spitzoid malignant melanomas. The American Journal of Dermatopathology. 2012; 34(1):82. https://doi.org/10.1097/DAD.0b013e31823df1e2 PMID: 22197864

17. Casadonte R, Kriegsmann M, Kriegsmann K, Streit H, Meliß RR, Müller CS, et al. Imaging Mass Spectrometry for the Classification of Melanoma Based on BRAF/NRAS Mutational Status. International Journal of Molecular Sciences. 2023; 24(6):5110. https://doi.org/10.3390/ijms24065110 PMID: 36982192

18. Lazova R, Seeley EH, Kutzner H, Scolyer RA, Scott G, Cerroni L, et al. Imaging mass spectrometry assists in the classification of diagnostically challenging atypical Spitzoid neoplasms. Journal of the American Academy of Dermatology. 2016; 75(6):1176–1186. https://doi.org/10.1016/j.jaad.2016.07.007 PMID: 27502312

19. Wang Y, Cai J, Louie DC, Wang ZJ, Lee TK. Incorporating clinical knowledge with constrained classifier chain into a multimodal deep network for melanoma detection. Computers in Biology and Medicine. 2021; 137:104812. https://doi.org/10.1016/j.compbiomed.2021.104812 PMID: 34507158

20. Schneider L, Wies C, Krieghoff-Henning EI, Bucher TC, Utikal JS, Schadendorf D, et al. Multimodal integration of image, epigenetic and clinical data to predict BRAF mutation status in melanoma. European Journal of Cancer. 2023; 183:131–138. https://doi.org/10.1016/j.ejca.2023.01.021 PMID: 36854237

21. Tuck M, Grélard F, Blanc L, Desbenoit N. MALDI-MSI Towards Multimodal Imaging: Challenges and Perspectives. Front Chem. 2022; 10. https://doi.org/10.3389/fchem.2022.904688 PMID: 35615316

22. Prade VM, Sun N, Shen J, Feuchtinger A, Kunzke T, Buck A, et al. The synergism of spatial metabolomics and morphometry improves machine learning-based renal tumour subtype classification. Clin Transl Med. 2022; 12(2). https://doi.org/10.1002/ctm2.666 PMID: 35184396

23. Erlmeier F, Feuchtinger A, Borgmann D, Rudelius M, Autenrieth M, Walch AK, et al. Supremacy of modern morphometry in typing renal oncocytoma and malignant look-alikes. Histochem Cell Biol. 2015; 144 (2):147–156. https://doi.org/10.1007/s00418-015-1324-4 PMID: 25929744

24. Chen T, Kornblith S, Norouzi M, Hinton G. A simple framework for contrastive learning of visual representations. In: International conference on machine learning. PMLR; 2020. p. 1597–1607.

25. Ciga O, Xu T, Martel AL. Self supervised contrastive learning for digital histopathology. Mach Learn. 2022; 7:100198.

26. Verbeeck N, Caprioli RM, Van de Plas R. Unsupervised machine learning for exploratory data analysis in imaging mass spectrometry. Mass Spectrom Rev. 2020; 39(3):245–291. https://doi.org/10.1002/mas.21602 PMID: 31602691

27. Liu X, Zhang F, Hou Z, Mian L, Wang Z, Zhang J, et al. Self-supervised learning: Generative or contrastive. IEEE Trans Knowl Data Eng. 2021;. https://doi.org/10.1109/TKDE.2021.3090866

28. Norris JL, Tsui T, Gutierrez DB, Caprioli RM. Pathology interface for the molecular analysis of tissue by mass spectrometry. J Pathol Inform. 2016; 7(1):13. https://doi.org/10.4103/2153-3539.179903 PMID: 27141319

29. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. Ieee; 2009. p. 248–255.

30. McInnes L, Healy J, Melville J. Umap: Uniform manifold approximation and projection for dimension reduction. arXiv. 2018;.

31. Smets T, Verbeeck N, Claesen M, Asperger A, Griffioen G, Tousseyn T, et al. Evaluation of distance metrics and spatial autocorrelation in uniform manifold approximation and projection applied to mass spectrometry imaging data. Anal Chem. 2019; 91(9):5706–5714. https://doi.org/10.1021/acs.analchem.8b05827 PMID: 30986042

32. Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995; 20(3):273–297. https://doi.org/10.1007/BF00994018

33. Luts J, Ojeda F, Van de Plas R, De Moor B, Van Huffel S, Suykens JA. A tutorial on support vector machine-based methods for classification problems in chemometrics. Anal Chim Acta. 2010; 665 (2):129–145. https://doi.org/10.1016/j.aca.2010.03.030 PMID: 20417323

34. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res. 2011; 12:2825–2830.

35. McInnes L, Healy J, Saul N, Grossberger L. UMAP: Uniform Manifold Approximation and Projection. J Open Source Softw. 2018; 3(29):861. https://doi.org/10.21105/joss.00861

36. Deininger SO, Bollwein C, Casadonte R, Wandernoth P, Gonçalves JPL, Kriegsmann K, et al. Multicenter Evaluation of Tissue Classification by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Imaging. Anal Chem. 2022;. https://doi.org/10.1021/acs.analchem.2c00097 PMID: 35658398

37. Gonçalves JPL, Bollwein C, Schlitter AM, Martin B, Märkl B, Utpatel K, et al. The impact of histological annotations for accurate tissue classification using mass spectrometry imaging. Metabolites. 2021; 11 (11):752. https://doi.org/10.3390/metabo11110752 PMID: 34822410

38. Reddy S. Explainability and artificial intelligence in medicine. The Lancet Digital Health. 2022; 4(4): e214–e215. https://doi.org/10.1016/S2589-7500(22)00029-2 PMID: 35337639

39. Amann J, Blasimme A, Vayena E, Frey D, Madai VI, Consortium P. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC medical informatics and decision making. 2020; 20:1–9. https://doi.org/10.1186/s12911-020-01332-6 PMID: 33256715

40. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision; 2017. p. 618–626.

41. Available from: https://paige.ai/. Date of access: 2023-12-22.