Finding reduced Raman spectroscopy fingerprint of skin samples for melanoma diagnosis through machine learning

Daniella Castro Araújo [a,b,*], Adriano Alonso Veloso [a], Renato Santos de Oliveira Filho [c], Marie-Noelle Giraud [d], Leandro José Raniero [e], Lydia Masako Ferreira [c], Renata Andrade Bitar [f,**]

[a] Computer Science Dept, Federal University of Minas Gerais, Brazil
[b] Kunumi, Brazil
[c] Dept of Surgery, Division of Plastic Surgery, Federal University of Sao Paulo, Brazil
[d] Dept of Medicine, University of Fribourg, Switzerland
[e] Research and Development Dept, Vale do Para'iba University, Brazil
[f] DERMiSCAN, Brazil

ABSTRACT

Early-stage detection of cutaneous melanoma can vastly increase the chances of cure. Excision biopsy followed by histological examination is considered the gold standard for diagnosing the disease, but requires long high-cost processing time, and may be biased, as it involves qualitative assessment by a professional. In this paper, we present a new machine learning approach using raw data for skin Raman spectra as input. The approach is highly efficient for classifying benign versus malignant skin lesions (AUC 0.98, 95% CI 0.97–0.99). Furthermore, we present a high-performance model (AUC 0.97, 95% CI 0.95–0.98) using a miniaturized spectral range (896–1039 cm^{-1}), thus demonstrating that only a single fragment of the biological fingerprint Raman region is needed for producing an accurate diagnosis. These findings could favor the future development of a cheaper and dedicated Raman spectrometer for fast and accurate cancer diagnosis.

1. Introduction

Excisional biopsy, followed by histopathological examination, is considered the gold standard for diagnosing melanoma and non-melanoma skin cancer [1,2]. However, removing each suspected skin lesion seems unacceptable, especially in cases of a giant or multiple lesions, mainly when located in aesthetically essential regions such as the face.

Currently, as a reference standard for the diagnosis of melanoma, clinical examination of skin lesions by a dermatologist is followed by biopsy and histopathological analysis. The difference in technical knowledge or experience of the evaluator exposes chronic cases of different diagnoses. The clinical exam sensitivity ranges from 40% to 80% of accurate diagnoses, depending on the physician's level of expertise [3–5]. Diagnosing malignant melanoma can be difficult due to its resemblance to specific benign lesions of the skin, such as pigmented nevi, and other types of skin cancers, such as basal cell carcinoma [6,7].

Braun et al. tested the diagnostic concordance among dermatopathologists in 1250 samples and revealed Kappa value of 0.8 for the final report between melanoma versus non-melanoma lesions and 0.62 for malignant melanocytic lesions versus borderline versus benign lesions [8].

An online tool created by the Global Cancer Observatory for the World Health Organization gives insight into how incidence and mortality rates for melanoma skin cancer could rise even further over time. This tool predicts that by 2025, the number of cases of melanoma skin cancer worldwide will rise by 18%, reaching 340,271, while the number of deaths is expected to rise 20%, reaching 72,886. By 2040, nearly half a million (466,914) people will be diagnosed with melanoma, an increase of 62%; while 105,904 will die from the disease, an increase of 74%. Furthermore, it shows that delaying treatment of a stage I melanoma by just one month increases the risk of death by 5% [9,10]. It's also important to note that the difference between the diagnostic and treatment processes for melanoma worldwide is striking. More developed countries are more likely to diagnose and treat quickly, whereas in less

---

developed countries people may not immediately seek medical treatment. It may be due to low income, lack of awareness of skin cancer signs, or only a lack of healthcare provision [11].

Even though many attempts to implement instrumental techniques as a reliable non-invasive methodology for detecting skin cancer (such as photographs, dermatoscopy, high-frequency ultrasound of the skin, and confocal microscopy in vivo), an accurate, quick, painless, and affordable device that meets the needs of patients and that is easy to adopt for a hospital's diagnostic routine has not yet been created [12–15]. Therefore, the creation of a disruptive and innovative methodology for clinical evaluation and support of non- or minimally invasive therapeutic indications for incipient lesions is of paramount importance for the accurate and preventive detection of melanoma and the clinical control of dysplastic lesions, avoiding multiple and expensive biopsies.

For this purpose, the development in biotechnology has intensified research on auxiliary instrumentation for the physical-chemical analysis of tissues, aiming to investigate the structures and interactions of biomolecular systems. In the perspective of acquiring biological data in a non-invasive way, Raman Spectroscopy has emerged as a technological alternative, because it allows access to information on the molecular structure of an investigated sample, and has been used for more than 70 years in physical-chemical non-destructive analyzes [16,17]. This method has been used by several researchers for detecting cutaneous, genital, brain, breast, and laryngeal tumors, revealing that the transition from healthy tissue to cancer is significantly associated with differences in the biochemical structures reflected in the Raman spectrum [17–21].

Several authors have highlighted the potential of Raman spectroscopy in disease clinical diagnosis [21–24]. Raman offers many benefits because it can be performed in a non- or minimally invasive way, and analysis can be rapid and accurate, which may lead to earlier interventions and therapeutic strategies that can reduce morbidity and mortality. Raman spectroscopy has been explored as a digital alternative for automated early diagnosis, with high precision and accuracy figures, easy to perform and that allows the evaluation of skin lesions in minutes [25], while the traditional analytical technique (biopsy and histological exams) requires hours or even days [26]. The pioneers of skin Raman spectroscopy have published the Raman spectral signatures of skin's stratum corneum and its annexes obtained in vitro. This data confirmed that Raman spectroscopy is a promising tool for investigating the molecular nature of the skin, because the Raman approach can study structural changes linked to various diseases [22]. As a high-precision molecular analytical technique, Raman spectroscopy can differentiate incipient lesions [27], making it possible to reduce the time or even the number of tests necessary to ensure the correct diagnosis [28].

Despite the uncomplicated handling of a Raman spectrometer for data acquisition experimentally, analysis of the spectra requires qualified expertise, for example, from specialists in Artificial Intelligence [29,30]. The desired accuracy for the classification of benign and malignant tissue Raman spectra requires a deeply structured analytical approach applied to the objectives of the studies. Avant-garde projects propose the application of state-of-the-art techniques in the Machine Learning area with the primary objective of enabling the classification of spectra with maximum sensitivity and specificity.

Santos et al. developed a diagnostic model by Raman Spectroscopy of cutaneous melanoma, validated in a set of independent data from 174 lesions. The diagnostic model correctly classified all melanomas with a specificity of 43.8% and sensitivity of 100%, which may indicate that it favors an approach to early clinical diagnosis of melanoma by Raman spectroscopy [31]. Aubertin et al. analyzed 947 Raman spectra of prostatic tissue using machine learning classification with neural networks and distinguished between benign and malignant spectra with a sensitivity of 87% and a specificity of 86% [32]. Khan et al. performed an analysis of hepatitis B virus (HBV) infection in human blood serum using Raman spectroscopy combined with a machine learning technique. Applying a support vector machine (SVM) algorithm with a polynomial kernel of order-2, they achieved an accuracy of 98%, sensitivity of 100%, and 95% of specificity in the classification of normal blood sera from HBV infected sera [33]. Dubey et al. used an SVM in the classification between Raman spectra of normal versus malignant breast tissue, obtaining an accuracy of 91%, sensitivity of 88%, and specificity of 93% [34].

Nonetheless, before applying any classification methods, handling the Raman spectra requires a deep understanding about its biological sample origin, the Raman technique used for the experiments, and the mathematical approaches to better extract Raman spectra statistical features. Firstly, fluorescence can be the most important disadvantage of Raman spectroscopy, especially for biological sample investigations [35,36]. Usually, the most important sources of autofluorescence in tissue samples are hemoglobin, melanin, collagen, lipids, and proteins [37]. Fluorescence intensity could be as great as 10^4 compared to the Raman signal [13,14]. This means that Raman peaks are overshadowed by a strong and broad fluorescence spectrum.

Therefore, when analyzing pigmented skin lesions, the influence of melanin fluorescence in the Raman of the skin data must be considered. Several studies have proposed automatic algorithms to remove fluorescence's influence on the Raman spectra, but none of these methods were able to resolve the broad melanin Raman bands, which usually are superimposed to the intrinsic fluorescence signal [38,39]. These spectral pre-processing methods used in most studies to date have attempted to remove background fluorescence and filter noise from pigmented skin Raman data, without considering that such information may be relevant for a more complete characterization of the structure. Most researchers agree that these procedures should be conducted with the utmost rigor, and should have no human intervention before the statistical analysis of the Raman spectra is complete [40].

In order to solve pre-processing issues, many studies have used statistical features to recognize patterns in time series. Lambrou and Kudumakis [41] use mean, variance, skewness, kurtosis, and entropy as statistical features to classify audio signals. Arnaout et al. use visual analytics techniques to explore some statistical features of sensor measurement [42]. Esmael et al. [43] calculate many statistical features to measure the properties of drilling data, like variance, mean, and skewness. Their results show how the statistical features are essential in detecting different situations in the underlying drilling process. Afseth et al. [44] attempted to robustly extract information from the various algorithms that are used to pre-process the Raman spectra of different samples. Their conclusions and recommendations influenced the study protocol to better deal with the background fluorescence also considering pigmented Raman data.

Accordingly, this present study has processed individual raw human skin Raman spectra to extract its critical statistical features. This paper's central goal is to propose a new machine learning approach for modeling Raman data, allowing biological sample identification through a specific miniaturized spectroscopy range. We intend to extract statistical features from pigmented skin Raman spectra raw data by identifying their most predictive mathematical features and representative spectroscopy range to distinguish melanoma from pigmented nevi. In summary, the main contributions of this paper are:

- We present a new approach of feature extraction for human skin Raman spectral data, focusing on their spectra's local properties. We fragmented each spectral range into continuous small subsets and calculated each element's statistical features;
- By employing the statistical features mentioned above as training data, we developed machine learning models capable of classifying Raman skin human tissue spectra among malignant Melanoma (ME) and benign Melanocytic Nevus (MN) with high performance (AUC 0.98, 95% CI 0.97–0.99), by analyzing the well-known "Raman biological fingerprint region" (800–1800 cm^{-1});
- Finally, and most importantly, we developed a high-performance model (AUC 0.97, 95% CI 0.95–0.98) using a miniaturized spectral

range (896–1039 cm^{-1}), proving that only a single and reduced fragment of the biological fingerprint Raman region is needed to classify benign versus malignant skin lesions. This finding adds new information to allow the future development of a cheaper and dedicated Raman spectrometer for cancer diagnosis.

2. Data and method

The study was consistent with the Ethical Principles for research involving humans as provided in Resolution 196/96 of the Brazilian National Health Council. All skin samples were obtained under a protocol approved by UNIFESP Institutional Review Board (Protocol 1895/07).

2.1. Dataset

For this study, 33 MN samples and 51 ME samples were obtained from surgical resection, stored at −196 °C, brought to room temperature in 0.9% saline solution, standardized in 2 mm^3 cuts, and further placed in a specific experimental sample holder. The FT-Raman spectrometer (RFS 100, Bruker®) was used to acquire data in vitro. An Nd:YAG laser @1064 nm excited the samples, and the Raman signals were collected by a Germanium detector cooled by liquid nitrogen, with the following acquisition parameters: laser power at 100–300 mW; 500 scans, and 2 cm^{-1} resolution. From each sample, five to ten spectra were acquired, at different points, in a "cross shape", separated by 100 μm increments in the x and y directions automatically. Such increments were adjusted by a micro-translation stage attached to the FT-Raman spectrometer. After spectra collection, all fragments were accordingly prepared to be sent for histopathological testing. Altogether, 436 Raman spectra were collected, 168 from MN, and 268 from ME.

The process of a differentiated and individualized skin lesions medical diagnosis considers both the patient's clinical and epidemiological context and the adequacy of these lesions to their morphological categories. Given the limited number of skin response patterns in the face of different pathological processes, the pattern-based classification method is a guide for the diagnosis of an unknown case. For pigmented lesions, one can consider grouping the following morphological categories into four differential diagnoses: (a) superficial and deep melanocytic nevus;

(b) pigment synthesizing dermal neoplasms; (c) primary tumorigenic melanomas; and (d) metastatic malignant melanoma [45]. In this study, MN (n = 168) and ME (n = 268) spectra were grouped, following the basic morphological categories criterion used for differential diagnosis. The MN group was composed of superficial and deep melanocytic nevus, and the ME group is composed of primary melanoma (n = 144) and metastatic melanoma (n = 124) spectra.

Fig. 1 shows the normalized mean intensity and standard deviation areas around the curves of the FT-Raman spectra of ME and MN groups in the 800 to 1800 cm^{-1} "biological fingerprint" spectral region.

"Biological Fingerprint" spectral features, commonly found in skin Raman spectra in the literature, were also observed similarly in our findings. Major differences between benign and malignant spectra were found in protein bands amide I (1640–1680 cm^{-1}), amide III (1220–1300 cm^{-1}), and n(C—C) stretching (probably in amino acids proline, hydroxyproline, 928–940 cm^{-1}), and in bands characteristic of lipids, CH2 scissoring vibration (1420–1450 cm^{-1}), and -(CH2)n- in-phase twist vibration around 1300 cm^{-1}; moreover, possible changes in polysaccharide structure were found in the region 840–860 cm^{-1} [22,23,38,39,44,46].

2.2. Fluorescence background data pre-process

Normally, most molecules exist in the ground electronic state (especially the lowest vibrational levels) at room temperature. After interacting with electromagnetic radiation, some molecules gain energy and occupy higher vibrational levels of the excited electronic states. After this, the molecules immediately lose their energy by collisions and go to the lowest vibrational level of the excited state. Finally, the molecules may fall to any of the vibrational levels of the ground state, so the energy differences between both electronic states arise as a form of fluorescence radiation [47].

In order to avoid fluorescence in the first place, choosing low laser frequencies as excitation sources could diminish the probability of fluorescence emission [17]. So, the use of higher wavelengths in near-infrared regions is helpful for reducing the fluorescence effect. Our skin data were collected employing the FT-Raman system equipped with a laser at 1064 nm, and in this excitation wavelength, the intrinsic fluorescence background level is extremely less pronounced in the
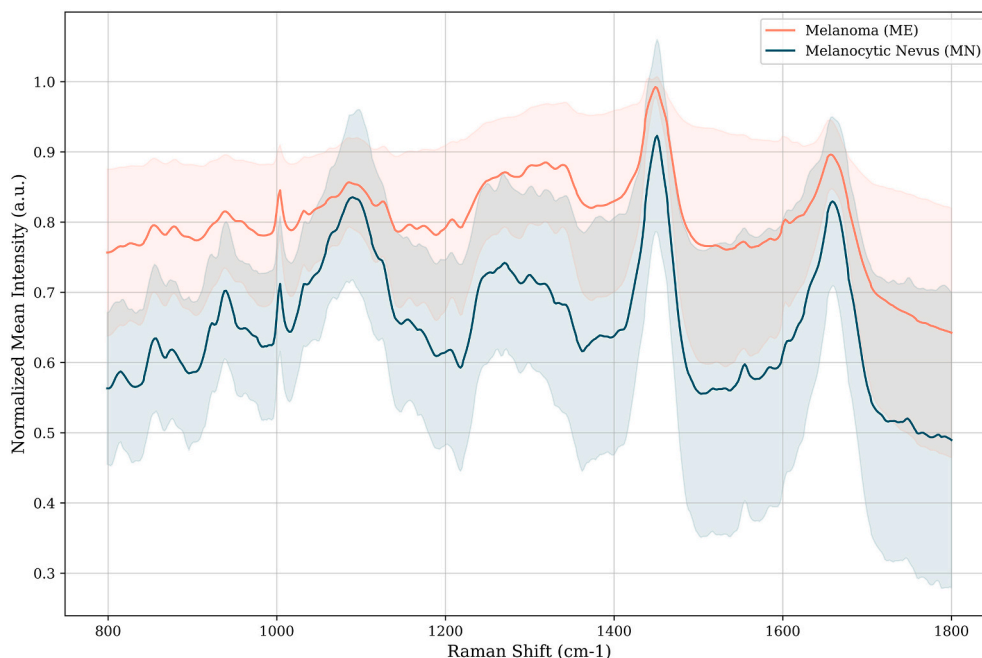

Fig. 1. Normalized mean intensity and standard deviation areas around the curves of the FT-Raman spectra of Melanoma (ME) and Melanocytic Nevus (MN) spectra.

region from 800 to 1800 cm^{-1} [38,47]. For that reason, we decided not to pre-process the data for autofluorescence, thus directly applying our new machine learning methodology, which involves extracting relevant statistical features from the raw data at the "biological fingerprint" region range [38,44].

Time-series data often exhibit inter-class differences in terms of small subsequences rather than the full series structure. Due to the similarity between the FT-Raman spectra and the general time-series data, part of the Ye and Keogh Shapelets technique was adapted to our machine learning protocol. Shapelets are defined as subsequences that are representative of a class. In a binary classification setting, a shapelet is discriminant if it is present in most series of one class and absent from the series of the other class. Using shapelets in a classification problem benefits models' explainability, speed, and accuracy [48].

2.3. Definitions and feature extraction

Definition 1. A *time series* $T \in \mathrm{IR}^n$ is a sequence of real-valued numbers $t_1 \in \mathrm{IR} : T = [t_1, t_2, ..., t_n]$, where n is the length of T. Similarly, a *spectrum* $S \in \mathrm{IR}^n$ is a sequence of real-valued numbers $s_1 \in \mathrm{IR}$, where i represents the spectral region in cm^{-1}. Thus, in our problem, $S = [s_{800}, ..., s_{1800}]$.

We are not interested in the global, but the local properties of a spectrum. A local region of a spectrum is called a *subsequence*.

Definition 2. A *subsequence* $S_{i, m}$ of a spectrum S is a continuous subset of the values from S of length m starting from spectral region i. Formally, $S_{i, m} = [s_i, s_{i+r}, ..., s_{i+(m-1)*r}]$, where r is the resolution acquisition parameter, and m is the defined subsequence length.

In order to extract relevant statistical features from the FT-Raman spectra, we performed the following two steps:

1. Sliced each spectrum S into subsequences $S_{i, m}$, each one containing the same length m. In order to select the appropriate granularity for the data, we tested different m's, varying from 5 to 100 contiguous data points.

2. For each subsequence $S_{i, m}$, we calculated seven statistical features shown by Table 1.

3. Results and discussion

3.1. Experimental setup — models construction and evaluation

The statistical features calculated were then used as input for the classification of ME vs. MN, using the fast LightGBM implementation [49]. LightGBM produces a complex model composed of hundreds of simple decision trees that are finally combined into a single model by a process known as boosting [50]. Random Forest, K-Nearest Neighbors (KNN), and XGBOOST algorithms were also tested, but, since their performance was very similar, only slightly worse than that of LightGBM, all the experiments reported in this work use LightGBM. To evaluate the performance of the models we used the standard area under the ROC curve (AUC) measure [51]. In order to improve the robustness of the estimations, we conducted five-fold cross-validation, that is, data is arranged into five folds, and at each run, four folds are used as a training set, and the remaining fold is used as a test set. We report the average AUC value over the five runs.

As a baseline, we averaged AUC values of models built using as input all the statistical features calculated for different subsequence lengths ($m = [5,25,50,100]$) in the entire spectral range (800–1800 cm^{-1}). Table 2 presents the average values of AUC and corresponding 95% Confidence Interval (95% CI) obtained for each m. We can observe that the AUC values do not significantly vary across different values of m.

In order to assess feature importance and thus extract intuitive insights from the prediction, we applied the SHAP algorithm [52] to every model. Briefly, SHAP calculates the importance of each statistical feature by estimating the effect of its absence on the model's decision. We plot, for each model, the five most important statistical features for every spectrum, and these results are depicted in the summary plot shown in Fig. 2. Red points are associated with spectra for which the corresponding statistical feature shows a relatively high value. Blue points, on the other hand, are associated with spectra for which the corresponding statistical feature shows a relatively low value. Further, there is a vertical line separating spectra associated with either negative (points on the left) or positive decisions (points on the right). For instance, points located in the left side are those for which the model provided a negative decision, that is, benign pigmented skin MN.

The most important feature observed in all the models is the Derivative. In order to evaluate simpler models – that use only this statistical feature – we performed experiments for each m, as can be seen in Table 3.

As we can see in Table 3, the AUC values of the models that use only the Derivatives are very similar to the models that use all the statistical features for $m = [5,25,50]$, being a little better for $m = 5$ and a little worse for $m = 25$ and $m = 50$. Considering $m = 100$, the AUC was 0.032 lower using only the Derivatives. Because (1) we are interested in creating models that use miniaturized spectroscopy regions, and (2) there was not a significant difference between different lengths (m's), from now on we are going to consider only models for $m = 5$ in the experiments. Besides, because (3) there was not a significant difference between the model that uses only Derivatives and the one that uses all the statistical features, we are going to use only the Derivatives as features in the experiments.

Table 1
Statistical features calculated for each subsequence $S_{i, m}$.

Statistical feature	Description	Formula
Arithmetic mean ($\mu_{i, m}$)	$\mu_{i, m}$ is the simple average of the values of a subsequence $S_{i, m}$.	$\mu_{i,m} = \frac{1}{m} \sum_{k=0}^{m-1} s_{i+kr}$
Standard deviation ($\sigma_{i, m}$)	$\sigma_{i, m}$ measures how the values of subsequence $S_{i, m}$ are spread out.	$\sigma_{i,m} = \sqrt{\frac{1}{m} \sum_{k=0}^{m-1} \overline{s_{i+kr} - \mu_{i,m}}^2}$
Kurtosis ($Ku_{i, m}$)	$Ku_{i, m}$ measures the peakedness of the probability distribution of subsequence $S_{i, m}$.	$Ku_{i,m} = \frac{\frac{1}{m} \sum_{k=0}^{m-1} (s_{i+kr} - \mu_{i,m})^4}{\sigma_{i,m}^4}$
Skewness ($Sk_{i, m}$)	$Sk_{i, m}$ is a measure of symmetry of subsequence $S_{i, m}$. A distribution is symmetric if it looks the same to the left and right of the corner point.	$Sk_{i,m} = \frac{\frac{1}{m} \sum_{k=0}^{m-1} (s_{i+kr} - \mu_{i,m})^3}{\sigma_{i,m}^3}$
Derivative ($D_{i, m}$)	$D_{i, m}$ measures the steepness of subsequence $S_{i, m}$.	$D_{i,m} = \frac{\mu_{i,m} - \mu_{(i-m*r),m}}{\mu_{i,m}}$
Maximum ($Max_{i, m}$)	$Max_{i, m}$ measures the maximum data point of subsequence $S_{i, m}$.	$Max_{i, m} = max (S_{i, m})$
Minimum ($Min_{i, m}$)	$Min_{i, m}$ measures the minimum data point of subsequence $S_{i, m}$.	$Min_{i, m} = min (S_{i, m})$

Table 2
Average AUC values and 95% CI range obtained using all the statistical features in the full spectral range (800–1800 cm^{-1}), separated by subsequence lengths (m).

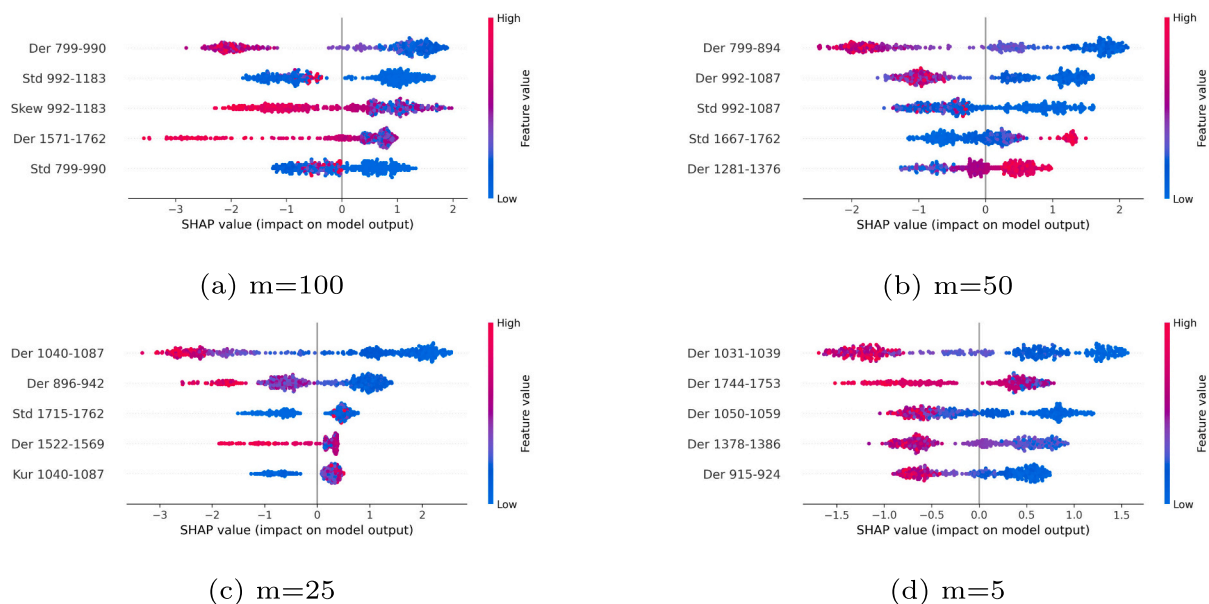| Subsequence length (m)			
5	25	50	100
0.978 (0.972–0.994)	0.980 (0.964–0.990)	0.988 (0.984–0.992)	0.982 (0.963–0.996)

(a) m=100

(b) m=50

(c) m=25

(d) m=5

Fig. 2. Summary plots showing the five most important statistical features for every model. The name of each feature in the plots is composed by the abbreviation of the Statistical Feature name plus the respective spectroscopy range — for example, 'Der 992–1183' is the Derivative calculated in the range 992–1183.

Table 3
Average AUC values and 95% CI range obtained using only Derivatives in the full spectral range (800–1800 cm^{-1}), separated by subsequence lengths (*m*).

Subsequence length (*m*)			
5	25	50	100
0.980	0.974	0.973	0.950
(0.965–0.991)	(0.962–0.991)	(0.963–0.995)	(0.928–0.969)

3.2. Finding an original predictive spectral range

In this subsection we present results for the classification of ME vs. MN using machine learning models created with parts of the total spectral range (800–1800 cm^{-1}). Our main goal is to identify an original predictive spectroscopy range for this classification, which could allow future miniaturization of the specified spectroscopy range for identifying benign vs malignant pigmented skin samples.

Finding the optimal machine learning model, i.e., the subset of features for which we achieve the best prediction accuracy, would require the exhaustive enumeration of all combinations of features. Alternatively, we sample the model space by randomly selecting an initial feature and a range size, and further randomly adding or not the next features within the chosen range size. At the end, we select the best performing models.

We sampled over thousands of mini-models - that consider spectral sizes varying only from 20 cm^{-1} to 300 cm^{-1} — for the classification of ME vs. MN, where each mini-model is composed by the Derivatives calculated for the subsequences of length $m = 5$ in different spectroscopy ranges. The range size and respective Derivatives within a model were randomly selected, thus resulting in models with diverse predictive performance. Of the thousands of mini-models generated, 158 achieved AUC higher than 0.960. Fig. 3 shows these best mini-models spectral ranges.

As we can see in Fig. 3, almost all of the best mini-models are concentrated in the spectral range of (800–1200 cm^{-1}).

3.3. Finding the best miniaturized model for differential diagnosis

As we can see in Fig. 4, seven mini-models achieved the highest AUC values, between 0.970 and 0.973. All of them use spectral regions
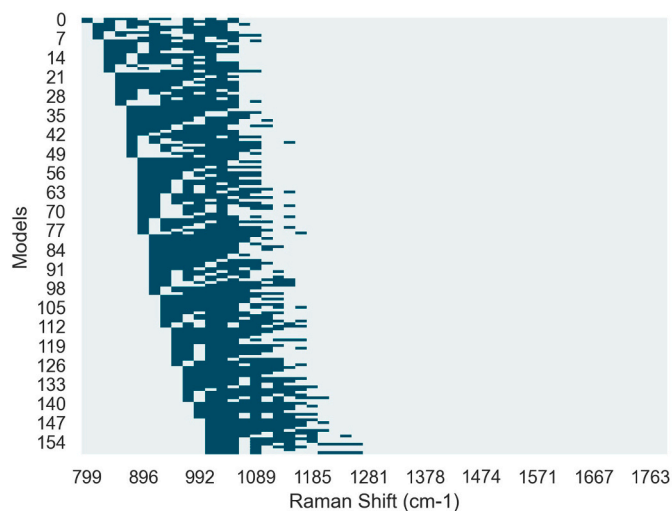

Fig. 3. The 158 mini-models that achieved AUC ≥0.960. Each line represents one model, and the dark green shows its spectral range. Empty spaces in the middle of some lines show that the Derivatives within that empty range were not used by the mini-model. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

between 838 and 1050 cm^{-1}, showing that this is the most predictive region for the separation of ME vs. MN.

Since our main objective in this study is to find a miniaturized Raman spectroscopy region that is able to differentiate ME from MN, we chose among the seven best mini-models the one with the smallest spectral size. This model is able to distinguish between benign and malignant pigmented skin spectra with an AUC of 0.973 95% CI 0.952–0.982), a sensitivity of 0.930 and a specificity of 0.870. Fig. 5 shows its ROC curve and corresponding AUC.

This model uses the Derivatives calculated in the region of 896–1039 cm^{-1}, thus using a total spectral size of only 143 cm^{-1}. Fig. 6 displays the SHAP Summary Plot of this mini-model, showing that the most important region for the model's decision making is the 915–924 cm^{-1}, followed by 1031–1039 cm^{-1}. These two regions are well known for C—C stretch of proline and hydroxyproline (collagen) amino acids,
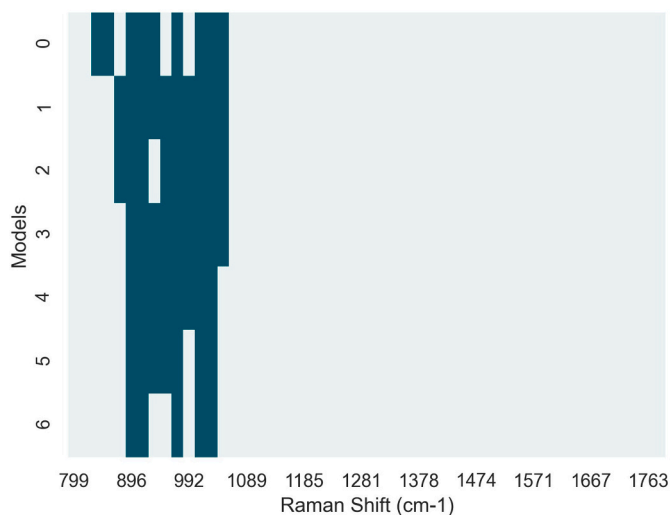
Fig. 4. The 7 mini-models that achieved AUC ≥0.970. Each line represents one model, and the dark green shows its spectral range. Empty spaces in the middle of some lines show that the Derivatives within that empty range were not used by the mini-model. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
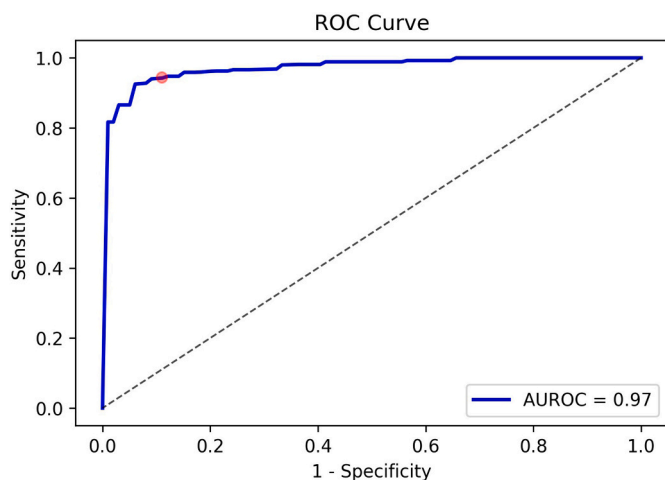

Fig. 5. ROC curve. The blue line shows the best miniaturized model predictions and the black dotted one shows random predictions. The orange dot shows the point where the sensitivity is 0.930 and the specificity is 0.870. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

glucose and lactic acid (extracellular-degrading enzymes), and C—H plane bending mode of phenylalanine (collagen), respectively [15,47,53].

Since we are working with melanin pigmented samples, it is essential to mention that melanin is not as relevant as collagen in differentiating melanoma from other pigmented skin lesions because melanomas do not always have more melanin than benign pigmented lesions [54]. Many studies also published some of the Raman spectral features of melanin in vitro [55,56] and in vivo [53,57]. It was found that free-melanin Raman bands D, D′, G, G′ were centered at 1347, 1407, 1554, and 1608 cm^{-1}, respectively, and attributed to stretching vibrations of the aromatic C=C bond. In conclusion, our seven mini-models that achieved the highest AUCs (between 0.970 and 0.973) employed the spectral regions between 838 and 1050 cm^{-1}, confirming that melanin Raman signals were unnecessary for the separation of ME vs. MN.

Feng et al. developed a biophysical basis and compared it to a classification model of Malignant Melanoma (MM) vs. Dysplastic Nevi (DN) Raman spectra. Their biophysical model revealed that the biomarkers responsible for the variances between these two pathologies were similar "fingerprint biological Raman band" reported here and in the literature [58–60]. Their best statistical analysis result indicated that collagen and triolein contributed significantly to MM and DN's spectral variance. AUC was 0.99, and specificity was 94% (90% to 95% sensitivity). They also found that the collagen content decreases significantly from benign lesions to cancer [54]. This decreased collagen in cancer was also observed in previous biophysical models of Raman ex vivo human skin fragments [61], breast tissue [21], gastric/esophagus tissue [62], and cervical tissue [62,63]. This may be partially explained by the release of metalloproteinases by cancerous cells to degrade dermal connective tissue [64,65] and extracellular-degrading enzymes secreted from fibroblasts that damage the stroma [66].

Consequently, we can correlate proline, hydroxyproline, phenylalanine, glucose, and lactic acid link with the collagen/extracellular matrix imbalance to the histopathological patterns in the melanoma progression disease features applied for a differential diagnosis: (a) dermal invasion, (b) inflammatory response at the tumor microenvironment, and (c) regression.

Dermal invasion is characterized by a proliferation of spindle-shaped, hyperchromatic melanocytes coursing in fascicles, nests, and single cells through the dermis. It usually progresses from an early radial growth phase confined within the epidermis to a vertical growth phase characterized by dermal invasion, where metastasis risk is high [66,67]. The complex inflammatory processes may support the understanding of the prognosis of melanoma. Different components of the tumor microenvironment could have stimulatory and inhibitory effects on melanoma progression by regulating its gene's expression repertoire within the tumor cells on the one hand and the stroma cells on the other, leading to the acquisition of the metastatic phenotype. This phenomenon is unusual in benign melanocytic proliferations [68]. Regression is also present in many invasive melanomas. It is a partial host response to a malignant neoplasm, resulting in the focal diminution of the process, and recently found to be a favorable prognostic factor for patients with initial stage melanomas. Histologically, the changes are similar to those seen in a scar [69].

Finally, the present paper compares some significant histopathological changes described in the Literature and the Raman spectra of skin pigmented lesions associating them to a possible melanoma differential diagnosis. This study aims to provide a Machine Learning perspective on the series of Raman biochemical changes that can be used to establish (or exclude) a diagnosis of melanoma. Fig. 7 shows a diagram with the methodology steps covered in this paper.

4. Conclusion

This paper studied the complexity of classifying raw skin Raman spectra as benign MN or malignant ME clinical conditions. Our proposed machine learning approach was built in four steps: (i) extraction of local statistical features in each spectrum; (ii) development of performant LightGsBM models (AUC ≥0.97) using the full available spectral "biological fingerprint" region (800–1800 cm^{-1}); (iii) identification of the most predictive spectral region, which turned out to be 800–1200 cm^{-1} and; (iv) development of a high performance model (AUC = 0.97) using the smallest possible region (896–1039 cm^{-1}) in which the most crucial region for the model's decision making was 915–924 cm^{-1}, followed by 1031–1039 cm^{-1}. Biochemically, when comparing these findings to a well-established Raman table data, these two smallest regions are known for the C—C stretch of proline, hydroxyproline, glucose, and lactic acid, and the C—H plane bending mode of phenylalanine. When related to the melanoma progress, the biochemical changes associated with these biochemical components refer to changes in the stroma's structural collagen adjacent lesions. More significant stromal permeability allows cellular invasion to adjacent structures, lymphatic and
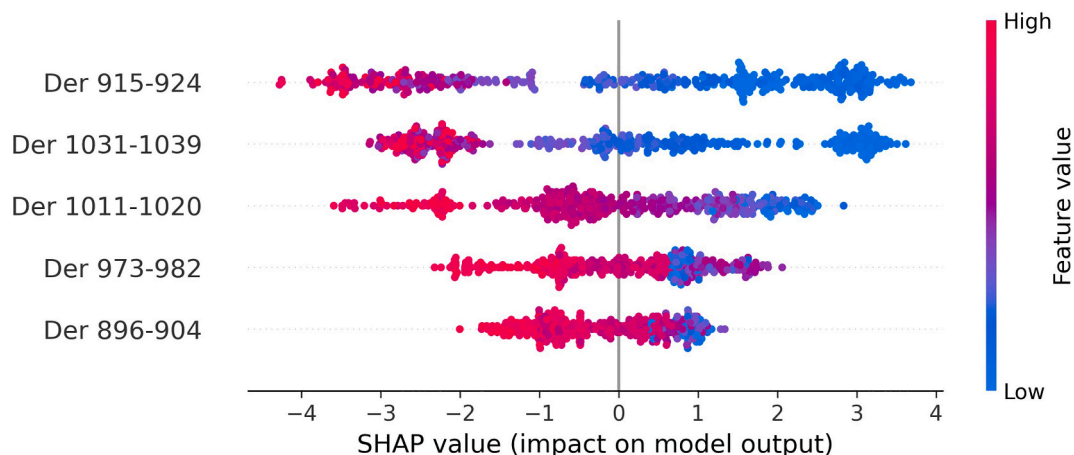
Fig. 6. Summary plot of the best mini-model generated (AUC = 0.973), showing the effects each Derivative has in identifying ME vs. MN spectra.

Raw Data

436 Spectra, being 168 Melanocytic Nevus (MN) and 268 Melanoma (ME)

Shapelets

We sliced each spectrum S into subsequences $S_{i,m}$, each one containing the length m. We tested different lengths: $m = [5, 25, 50, 100]$.

Feature Extraction

We extracted Statistical Features for each subsequence $S_{i,m}$ - Arithmetic mean, Standart devitation, Kurtosis, Skewness, Derivative, Maximum and Minimum.

Baseline

We estimated AUC values for models that used all the statistical features calculated for all legths m in the entire spectral range: 800 - 1800 cm^1, using 5-fold cross validation.

Mini-Models Exploration

Using m=5 and Derivatives as features, we have randomly generated thousands of mini-models - i.e., models that use only a part of the total spectral range.

Best Model Selection

We evaluated all the mini-models in terms of AUC, using 5-fold cross validation. We ended up with 7 performant mini-models (AUC = 0.97), and chose the one that used the smallest spectral range (896 - 1039 cm^{-1}).
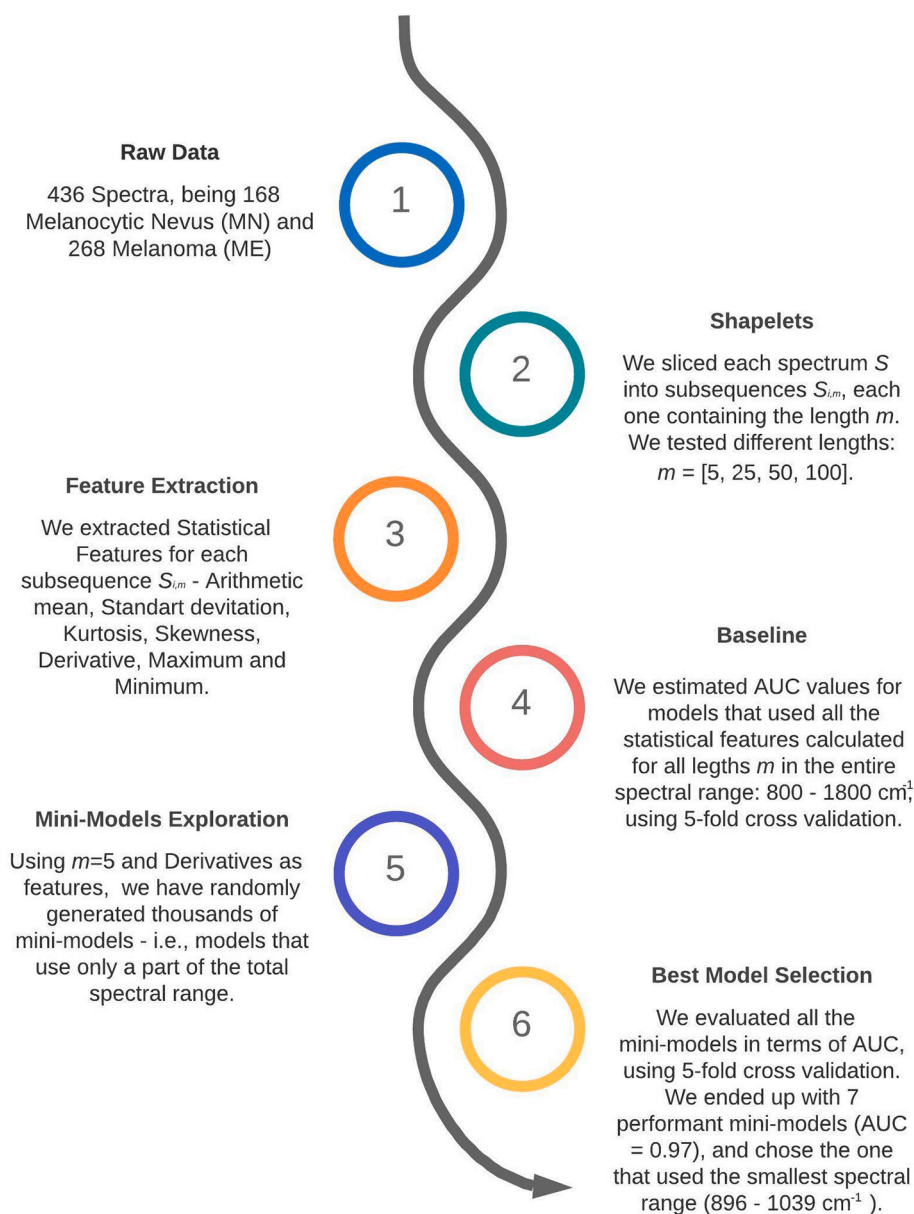
Fig. 7. Diagram summarizing the steps we have followed in this work.

blood vessels, and it may cause inflammatory infiltration reactions. In addition, melanin is not as relevant as collagen in differentiating melanoma from other pigmented skin lesions because melanomas do not always have more melanin than benign pigmented lesions, confirming that melanin Raman signals were unnecessary for the separation of ME vs. MN using our seven mini-models approach. Therefore, our machine learning approach on skin Raman raw data could allow the future development of a cheaper and dedicated spectrometer for a dermatological routine checkup. It may increase the early-stage detection of melanoma rate, also avoiding unnecessary biopsy and histopathological analysis, consequently reducing health costs to save lives.

It is worth mentioning that the evidence base shown so far is limited, as these device technologies and mathematical approaches have not yet been incorporated into a clinical routine. To date, it is not clear how the new Raman devices would perform in a non-research environment or what would be the ideal location for diagnosis. Multicenter confirmatory studies will be useful in this regard.

Declaration of competing interest

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

Acknowledgments

References

[1] Okun MR, Edelstein LM, Kasznica J. What criteria reliably distinguish melanoma from benign melanocytic lesions? Histopathology 2000;37(5):464–6. https://doi.org/10.1046/j.1365-2559.2000.01002.x.

[2] Slater DN. Doubt and uncertainty in the diagnosis of melanoma. Histopathology 2000;37(5):469–72. https://doi.org/10.1046/j.1365-2559.2000.10023.x.

[3] Cassileth BR, Clark Jr WH, Lusk EJ, Frederick BE, Thompson CJ, Walsh WP. How well do physicians recognize melanoma and other problem lesions? J Am Acad Dermatol 1986;14(4):555–60.

[4] Morton Mackie. Clinical accuracy of the diagnosis of cutaneous malignant melanoma. Br J Dermatol 1998;138(2):283–7. https://doi.org/10.1046/j.1365-2133.1998.02075.x.

[5] Chen SC, Bravata DM, Weil E, Olkin I. A comparison of dermatologists' and primary care physicians' accuracy in diagnosing melanoma: a systematic review. Arch Dermatol 2001;137(12):1627–34. https://doi.org/10.1001/archderm.137.12.1627.

[6] MacKie RM. Malignant melanoma: clinical variants and prognostic indicators. Clin Exp Dermatol 2000;25(6):471–5. https://doi.org/10.1046/j.1365-2230.2000.00692.x.

[7] Kanzler MH, Mraz-Gernhard S. Primary cutaneous malignant melanoma and its precursor lesions: diagnostic and therapeutic overview. J Am Acad Dermatol 2001;45(2):260–76. https://doi.org/10.1067/mjd.2001.116239.

[8] Braun RP, Gutkowicz-Krusin D, Rabinovitz H, Cognetta A, Hofmann-Wellenhof R, Ahlgrimm-Siess V, et al. Agreement of dermatopathologists in the evaluation of clinically difficult melanocytic lesions: how golden is the 'gold standard'? Dermatology 2012;224(1):51–8.

[9] Cancer facts and figures. https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2021/cancer-facts-and-figures-2021.pdf; 2021. accessed: 2021-3-2.

[10] 2020 melanoma skin cancer report stemming the global epidemic. https://www.melanomauk.org.uk/Handlers/Download.ashx?IDMF=91e70826-91d1-4b5e-9b0c-3dd3da10686d. accessed: 2021-3-2.

[11] Jaganathan C. Global green and bio solvents market, 2016–2020. Ind Biotechnol 2016;12:216–8. https://doi.org/10.1089/ind.2016.29044.cbj.

[12] Harland CC, Kale SG, Jackson P, Mortimer PS, Bamber JC. Differentiation of common benign pigmented skin lesions from melanoma by high-resolution ultrasound. Br J Dermatol 2000;143(2):281–9.

[13] Bafounta M-L, Beauchet A, Aegerter P, Saiag P. Is dermoscopy (epiluminescence microscopy) useful for the diagnosis of melanoma? Arch Dermatol 2001;137(10). https://doi.org/10.1001/archderm.137.10.1343.

[14] Kittler H, Menzies SW. Diagnostic accuracy of dermoscopy. In: Atlas of dermoscopy; 2012. p. 351–3.

[15] Prichard RS, Hill ADK, Skehan SJ, O'Higgins NJ. Positron emission tomography for staging and management of malignant melanoma. Br J Surg 2002;89(4):389–96.

[16] Edwards HGM, Williams AC, Barry BW. Potential applications of FT-Raman spectroscopy for dermatological diagnostics. J Mol Struct 1995;347:379–87.

[17] Hanlon EB, Manoharan R, Koo TW, Shafer KE, Motz JT, Fitzmaurice M, et al. Prospects for in vivo raman spectroscopy. Phys Med Biol 2000;45(2):R1–59. https://doi.org/10.1088/0031-9155/45/2/201.

[18] Liu CH, Das BB, Glassman WL Sha, Tang GC, Yoo KM, Zhu HR, et al. Raman, fluorescence, and time-resolved light scattering as optical diagnostic techniques to separate diseased and normal biomedical media. J Photochem Photobiol B 1992;16(2):187–209.

[19] Mizuno A, Kitajima H, Kawauchi K, Muraishi S, Ozaki Y. Near-infrared fourier transform raman spectroscopic study of human brain tissues and tumours. J Raman Spectr 1994;25(1):25–9.

[20] Stone N, Kendall C, Smith J, Crow P, Barr H. Raman spectroscopy for identification of epithelial cancers. Faraday Discuss 2004;126:141–57. https://doi.org/10.1039/B304992B.

[21] Bitar RA, S. Martinho Hd, Tierra-Criollo CJ, Ramalho LN Zambelli, Netto MM, Martin AA. Biochemical analysis of human breast tissues using Fourier-transform raman spectroscopy. J Biomed Opt 2006;11(5):054001.

[22] Gniadecka M, Wulf HC, Nielsen OF, Christensen DH, Hercogova J. Distinctive molecular abnormalities in benign and malignant skin lesions: studies by Raman spectroscopy. Photochem Photobiol 1997;66(4):418–23.

[23] Ellis DI, Goodacre R. Metabolic fingerprinting in disease diagnosis: biomedical applications of infrared and Raman spectroscopy. Analyst 2006;131(8):875–85.

[24] Lieber CA, Majumder SK, Billheimer D, Ellis DL, Mahadevan-Jansen A. Raman microspectroscopy for skin cancer detection in vitro. J Biomed Opt 2008;13(2):024013.

[25] S. A. Fox, A. A. Shanblatt, H. Beckman, J. Strasswimmer, A. C. Terentis, Raman spectroscopy differentiates squamous cell carcinoma (SCC) from normal skin following treatment with a high-powered CO_2 laser, Lasers Surg Med 46 (10) (2014) 757–772.

[26] Huang Z, McWilliams A, Lui H, McLean DI, Lam S, Zeng H. Near-infrared raman spectroscopy for optical diagnosis of lung cancer. Int J Cancer 2003;107(6):1047–52.

[27] Michalska M, Chodorowska G, Krasowska D. SIAscopy—a new non-invasive technique of melanoma diagnosis. Ann Univ Mariae Curie Sklodowska Med 2004;59(2):421–31.

[28] Leslie DG, Kast RE, Poulik JM, Rabah R, Sood S, Auner GW, et al. Identification of pediatric brain neoplasms using raman spectroscopy. Pediatr Neurosurg 2012;48(2):109–17.

[29] Zeng H, Zhao J, Short M, Mclean DI, Lam S, Mcwilliams A, et al. Raman spectroscopy for in vivo tissue analysis and diagnosis, from instrument development to clinical applications. J Innov Optic Health Sci 2008;01(01):95–106. https://doi.org/10.1142/S1793545808000054.

[30] Zhao J, Lui H, Kalia S, Zeng H. Real-time raman spectroscopy for automatic in vivo skin cancer detection: an independent validation. Anal Bioanal Chem 2015;407(27):8373–9.

[31] Santos IP, Doorn R van, Caspers PJ, Schut TC Bakker, Barroso EM, Nijsten TEC, et al. Improving clinical diagnosis of early-stage cutaneous melanoma based on Raman spectroscopy. Br J Cancer 2018;119(11):1339–46. https://doi.org/10.1038/s41416-018-0257-9.

[32] Aubertin K, Trinh VQ, Jermyn M, Baksic P, Grosset A-A, Desroches J, et al. Mesoscopic characterization of prostate cancer using raman spectroscopy: potential for diagnostics and therapeutics. BJU Int 2018;122(2):326–36. https://doi.org/10.1111/bju.14199.

[33] Khan S, Ullah R, Khan A, Ashraf R, Ali H, Bilal M, et al. Analysis of hepatitis b virus infection in blood sera using raman spectroscopy and machine learning. Photodiagnosis Photodyn Ther 2018;23:89–93. https://doi.org/10.1016/j.pdpdt.2018.05.010.

[34] Dubey V, Ahmad A, Butola A, Qaiser D, Srivastava A, Mehta DS. Low coherence quantitative phase microscopy with machine learning model and raman spectroscopy for the study of breast cancer cells and their classification. Appl Optics 2019;58(5):A112–9. https://doi.org/10.1364/AO.58.00A112.

[35] Shreve AP, Cherepy NJ, Mathies RA. Effective rejection of fluorescence interference in raman spectroscopy using a shifted excitation difference technique. Appl Spectrosc 1992;46(4):707–11. URL, http://as.osa.org/abstract.cfm?URI=as-46-4-707.

[36] Yang S, Li B, Slipchenko MN, Akkus A, Singer NG, Yeni YN, et al. Laser wavelength dependence of background fluorescence in raman spectroscopic analysis of synovial fluid from symptomatic joints. J Raman Spectr 2013;44(8):1089–95. https://doi.org/10.1002/jrs.4338.

[37] Harland C, Kale S, Jackson P, Mortimer P, Bamber J. Differentiation of common benign pigmented skin lesions from melanoma by high-resolution ultrasound. Br J Dermatol 2000;143(2):281–9. https://doi.org/10.1046/j.1365-2133.2000.03652.x.

[38] Gniadecka M, Philipsen PA, Sigurdsson S, Wessel S, Nielsen OF, Christensen DH, et al. Melanoma diagnosis by raman spectroscopy and neural networks: structure alterations in proteins and lipids in intact cancer tissue. J Invest Dermatol 2004;122(2):443–9.

[39] Sigurdsson S, Philipsen PA, Hansen LK, Larsen J, Gniadecka M, Wulf HC. Detection of skin cancer by classification of Raman spectra. IEEE Trans Biomed Eng 2004;51(10):1784–93.

[40] Cao A, Pandya AK, Serhatkulu GK, Weber RE, Dai H, Thakur JS, et al. A robust method for automated background subtraction of tissue fluorescence. J Raman Spectr 2007;38(9):1199–205. arXiv:.

[41] T. Lambrou, P. Kudumakis, R. Speller, M. Sandler, A. Linney, Classification of audio signals using statistical features on time and wavelet transform domains, Proceedings of the 1998 IEEE international conference on acoustics, speech and signal processing, ICASSP '98 (Cat. No. 98CH36181).

[42] Arnaout A, Alsallakh B, Fruhwirth R, Thonhauser G, Esmael B, Prohaska M. Diagnosing drilling problems using visual analytics of sensors measurements. In: 2012 IEEE international instrumentation and measurement technology conference proceedings; 2012.

[43] Arnaout A, O'Leary P, Esmael B, Thonhauser G. Distributed recognition system for drilling events detection and classification. Int J Hybr Intell Syst 2013;11(1): 25–39.

[44] Afseth NK, Segtnan VH, Wold JP. Raman spectra of biological samples: a study of preprocessing methods. Appl Spectrosc 2006;60(12):1358–67.

[45] Elder D. Atlas of dermatopathology: synopsis and atlas of lever's histopathology of the skin. LWW; 2020.

[46] Mendelsohn R, Flach CR, Moore DJ. Determination of molecular conformation and permeation in skin via IR spectroscopy, microscopy, and imaging. Biochim Biophys Acta 2006;1758(7):923–33.

[47] Kocademir M, Kumru M, Gölcük K, Suarez-Ibarrola R, Miernik A. Fluorescence reduction in raman spectroscopy by chemical bleaching on renal stones. J Appl Spectr 2020;87(2):282–8. https://doi.org/10.1007/s10812-020-00997-1.

[48] Ye L, Keogh E. Time series shapelets: a novel technique that allows accurate, interpretable and fast classification. Data Min Knowl Discov 2011;22(1–2):149–82.

[49] Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: a highly efficient gradient boosting decision tree. 2017. p. 3146–54.

[50] Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci 1997;55(1):119–39. https://doi.org/10.1006/jcss.1997.1504.

[51] Fawcett T. An introduction to roc analysis. Patt Recogn Lett 2006;27(8):861–74. rOC Analysis in Pattern Recognition, https://doi.org/10.1016/j.patrec.2005.10.010.

[52] Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. 2017. p. 4765–74.

[53] Huang Z, Zeng H, Hamzavi I, Alajlan A, Tan E, McLean DI, et al. Cutaneous melanin exhibiting fluorescence emission under near-infrared light excitation. J Biomed Opt 2006;11(3):34010.

[54] Feng X, Moy AJ, Nguyen HTM, Zhang Y, Zhang J, Fox MC, et al. Raman biophysical markers in skin cancer diagnosis. J Biomed Opt 2018;23(5):1–10.

[55] Capozzi V, Perna G, Gallone A, Biagi PF, Carmone P, Fratello A, et al. Raman and optical spectroscopy of eumelanin films. J Mol Struct 2005;744–747:717–21.

[56] Perna G, Gallone A, Capozzi V, Biagi PF, Fratello A, Guida G, et al. Optical spectra of melanin films extracted from Rana esculenta L. Phys Scr 2005:89.

[57] Huang Z, Lui H, Chen XK, Alajlan A, McLean DI, Zeng H. Raman spectroscopy of in vivo cutaneous melanin. J Biomed Opt 2004;9(6):1198–205.

[58] Feng X, Moy AJ, Nguyen HTM, Zhang J, Fox MC, Sebastian KR, et al. Raman active components of skin cancer. Biomed Opt Express 2017;8(6):2835–50.

[59] Pezzotti G, Boffelli M, Miyamori D, Uemura T, Marunaka Y, Zhu W, et al. Raman spectroscopy of human skin: looking for a quantitative algorithm to reliably estimate human age. J Biomed Opt 2015;20(6):065008.

[60] Barry BW, Edwards HGM, Williams AC. Fourier transform Raman and infrared vibrational study of human skin: assignment of spectral bands. J Raman Spectr 1992;23(11):641–5.

[61] Silveira Jr L, Silveira FL, Bodanese B, Zângaro RA, Pacheco MTT. Discriminating model for diagnosis of basal cell carcinoma and melanoma in vitro based on the Raman spectra of selected biochemicals. J Biomed Opt 2012;17(7):077003.

[62] Andrade PO, Bitar RA, Yassoyama K, Martinho H, Santo AME, Bruno PM, et al. Study of normal colorectal tissue by FT-Raman spectroscopy. Anal Bioanal Chem 2007;387(5):1643–8.

[63] Teixeira CSB, Bitar RA, Martinho HS, Santos ABO, Kulcsar MAV, Friguglietti CUM, et al. Thyroid tissue analysis through raman spectroscopy. Nov.. 2009.

[64] Zucker S, Turpeenniemi-Hujanen T, Ramamurthy N, Wieman J, Lysik R, Gorevic P, et al. Purification and characterization of a connective-tissue-degrading metalloproteinase from the cytosol of metastatic melanoma cells. Biochem J 1987; 245(2):429–37.

[65] Quan T, Qin Z, Xia W, Shao Y, Voorhees JJ, Fisher GJ. Matrix-degrading metalloproteinases in photoaging. J Investig Dermatol Symp Proc 2009;14(1): 20–4.

[66] Watt FM, Fujiwara H. Cell-extracellular matrix interactions in normal and diseased skin. Cold Spring Harb Perspect Biol 2011;3(4). Apr.

[67] Kharbili M El, Cario M, Béchetoille N, Pain C, Boucheix C, Degoul F, et al. Tspan8 drives melanoma dermal invasion by promoting ProMMP-9 activation and basement membrane proteolysis in a keratinocyte-dependent manner. Cancers 2020;12(5). May.

[68] Bar-Eli M. Regulation of gene expression in the tumor environment: regulation of melanoma progression by the microenvironment: the roles of PAR-1 and PAFR. Springer Science & Business Media; 2008.

[69] Sharouni M-A El, Aivazian K, Witkamp AJ, Sigurdsson V, van Gils CH, Scolyer RA, et al. Association of histologic regression with a favorable outcome in patients with stage 1 and stage 2 cutaneous melanoma. JAMA Dermatol 2021;157(2):166–73.