Portable System for In-Clinic Differentiation of Skin Cancers from Benign Skin Lesions and Inflammatory Dermatoses

Michel Nieuwoudt[1,2,3,4], Paul Jarrett[5,6], Hannah Matthews[1,2,3,4], Michelle Locke[7,8], Marco Bonesi[1,2,3,4,9], Brydon Burnett[1,2], Hannah Holtkamp[1,2,3], Claude Aguergaray[2,3,4,9], Ira Mautner[2,3,4,9], Thom Minnee[1,2,3,4] and M. Cather Simpson[1,2,3,4,9]

The exquisite sensitivity of Raman spectroscopy for detecting biomolecular changes in skin cancer has previously been explored; however, this mostly required analysis of excised tissue samples using bulky, immobile laboratory instrumentation. In this study, the technique was translated for clinical use with a portable Raman system and customized fiber optic probe and applied to differentiation of skin cancers from benign lesions and inflammatory dermatoses. The aim was to provide an easy-to-use, easy-to-manage assessment tool for clinicians to use in their daily patient examination routine to perform rapid Raman measurements of skin lesions in vivo. Using this system, >867 spectra were measured in vivo from 330 patients with a wide variety of different benign skin lesions (n = 603), inflammatory dermatoses (n = 140), and skin cancers (n = 124). Ethnicities represented were 70% European; 16% Asian; 6% Māori; 5% Pacific people; and 4% Middle East, Latin American, and African. Accurate differentiation of skin cancers from benign lesions and inflammatory dermatoses was achieved using partial least squares discriminant analysis, with area under curve for the receiver operator curves for external validation sets ranging from 0.916 to 0.958. This study shows evidence for robust clinical translation of Raman spectroscopy for rapid, accurate diagnosis of skin cancer.

INTRODUCTION

The appropriate use of modern technology such as teledermatology and artificial intelligence in diagnosing dermatological disease offers great potential for effective triage of skin cancer referral, particularly in resource-constrained environments. The clinician using these technologies to triage these referrals needs to be experienced and trained in the use of the technology and in recognizing the disease. The patient, referrer, and clinician do not want to miss a clinically significant lesion that may require urgent treatment, for example, a malignant melanoma. Cost-effective triage should reliably screen out benign lesions, permitting resources to focus on the treatment of malignant lesions. However, referrals to a skin cancer service often lack the specific detail needed for accurate triage and may be sent with limited history and poor-quality images from which an accurate triage assessment cannot be made. Furthermore, the remote but accurate diagnosis of an inflammatory dermatosis may expedite treatment when there are delays in seeing a specialist dermatologist. Access problems are amplified when there are additional geographic or financial barriers to expert care. The combination of clinical details combined with the application of in vivo Raman spectroscopy may offer further refinement to the triage process.

Raman spectroscopy uses laser radiation to measure the characteristic vibrational frequencies of molecules, providing a specific fingerprint of the individual compounds forming complex systems, such as human skin. When a probing laser is shone on to any biological sample, the light interacts with the molecules and undergoes changes in energy. The molecules then scatter the light, and roughly one millionth of the scattered photons experience frequency shifts (inelastic scattering) when interacting with the vibrating molecules, these frequencies that provide information on the types of molecules and their molecular environment. The inelastically scattered light is Raman scattering, and the intensities of the Raman scattering can be picked with charge-coupled device sensors in Raman spectrometers, which measure the intensities of the scattered light as a function of frequency. The Raman scatter is plotted as a Raman spectrum of peak intensities of scattered photons against wavenumbers (cm^{-1}) that represent the frequencies of the specific molecular

[1]School of Chemical Sciences, University of Auckland, Auckland, New Zealand; [2]The Photon Factory, University of Auckland, Auckland, New Zealand; [3]The Dodd Walls Centre for Photonic and Quantum Technologies, Dunedin, New Zealand; [4]The MacDiarmid Institute for Advanced Materials and Nanotechnology, Wellington, New Zealand; [5]Department of Dermatology, Middlemore Hospital, Auckland, New Zealand; [6]Department of Medicine, University of Auckland, Auckland, New Zealand; [7]Department of Plastic Surgery, Middlemore Hospital, Auckland, New Zealand; [8]Department of Surgery, University of Auckland, Auckland, New Zealand; and [9]Department of Physics, University of Auckland, Auckland, New Zealand

Correspondence: Michel Nieuwoudt, School of Chemical Sciences, University of Auckland, 23 Symonds Street, Auckland 1142, New Zealand. E-mail: m.nieuwoudt@auckland.ac.nz

vibrations in the sampled tissue. The set of frequencies is characteristic of specific molecules so the Raman spectrum provides a unique chemical fingerprint of the composition of the tissue. When low laser powers are used, the technique is noninvasive and nondestructive, thus highly suited for identifying differences in biomolecules in malignant and benign cells (Barroso et al, 2015; Brozek-Pluska et al, 2012; Cals et al, 2015; Carvalho et al, 2015; Huang et al, 2003; Kong et al, 2013; Santos et al, 2017; Stone et al, 2004).

The use of a Raman microscope with a Raman spectrometer allows for measurement of microsized areas of tissue and has been used to examine skin composition and hydration (Baclig et al 2013; Essendoubi et al 2016; Gniadecka et al, 2003, 1998; Nguyen et al 2012; Zhang et al 2011). It has also been investigated for classifying skin cancers, including basal cell carcinoma (BCC) (Choi et al, 2005; Feng et al, 2018; Larraona-Puy et al, 2009; Lieber et al, 2008a, 2008b; Nijssen et al, 2007; Silveira et al, 2015), squamous cell carcinoma (SCC) (Fox et al, 2014; Lieber et al, 2008a, 2008b; Lim et al, 2014; Silveira et al, 2015), and melanoma (Gniadecka et al, 2004; Hammody et al, 2005; P Santos et al, 2018; Lieber et al, 2008a; Lui et al, 2012; Santos et al, 2016). These studies have shown the potential of this technique as a diagnostic tool for early detection of melanoma and non-melanoma skin cancers (NMSCs). However, these studies lack useful clinical translation because they have been applied ex vivo on excised tissue samples in the laboratory using large and expensive equipment.

Several studies (Bratchenko et al, 2021; Feng et al, 2018; Khristoforova et al, 2019; Kourkoumelis et al, 2015; Lieber et al, 2008b; Lim et al, 2014; Lui et al, 2012; Sharma et al, 2014; Silveira et al, 2015) have investigated the feasibility of applying Raman spectroscopy as a diagnostic tool for in vivo measurement of skin cancers using smaller, portable benchtop spectrometers of various sizes. Excitation wavelengths in the near-infrared region between 785 and 1064 nm were selected to minimize the fluorescence exhibited by melanin in the skin because these effectively mask the Raman signals. In some of the studies, the autofluorescence signal was included in the measurement (Bratchenko et al, 2021; Khristoforova et al, 2019), and different fiber optic probes and classification methods were developed for some of the portable systems (Latka et al, 2013; Motz et al, 2004; Sharma et al, 2014). However, the studies were limited to skin neoplasms, and sample numbers were commonly small (range = 21−518); in addition, they almost exclusively included European patient samples. This study focused on developing a Raman spectroscopy—based portable device suitable for deployment in clinical practice that would allow primary clinicians to perform accurate in vivo, real-time differentiation of malignant skin lesions from a wider variety of lesions that also included a wider variety of ethnicities. Included in the study were melanomas; NMSCs; benign lesions; and inflammatory dermatoses from ethnic groups, including European, Maori and Pacific Island people, Asians and Latin American, African, and Middle Eastern. All biopsied skin lesions were confirmed as malignant or benign by histopathology, whereas additional benign lesions and inflammatory dermatoses that could not be biopsied were clinically diagnosed by a highly experienced dermatologist, following standard clinical practice.

A commercially available, portable Raman spectrometer (EmVision HT) with deep depletion cooled charge couple device detector and a custom-made fiber optic probe (EmVision, Tampa, FL) were used for in vivo measurements in the clinic (Figure 1).

The laser excitation wavelength of 830 nm was selected to reduce the autofluorescence by melanin exhibited for 785 nm lasers. Other Raman studies on skin cancers have used 785 nm excitation (Bratchenko et al, 2021, 2017; Lui et al, 2012) because this provides better Raman intensities overall in a shorter time and better detector efficiency for the higher wavenumber regions >1500 cm^{-1}. These studies also included the autofluorescence signal in the classification (Bratchenko et al, 2017; Khristoforova et al, 2019), which was found to afford better classification accuracy (Bratchenko et al, 2017). Skin measurements were optimized through benchtop testing to ensure that high-quality Raman spectra could be obtained from consented patients in the shortest time possible, while also ensuring conformity with the American National Standards Institute maximum permitted exposure for skin. Interaction time during measurement was minimized to allow seamless incorporation into the clinic patient flow. Classification of skin measurement was performed using partial least squares with discriminant analysis (PLS-DA) with random subset cross-validation of the training model, and the classification model was validated with a separate validation set. The PLS-DA algorithm was selected being the preferred method for practical classification of complex chemical data imbalanced datasets such as the one generated from our study. The method is very suitable for classifying Raman spectra of biological tissue and has also been used in previous Raman studies of skin cancers (Khristoforova et al, 2019; Lui et al, 2012). The variations in different malignant and benign skin lesions and inflammatory
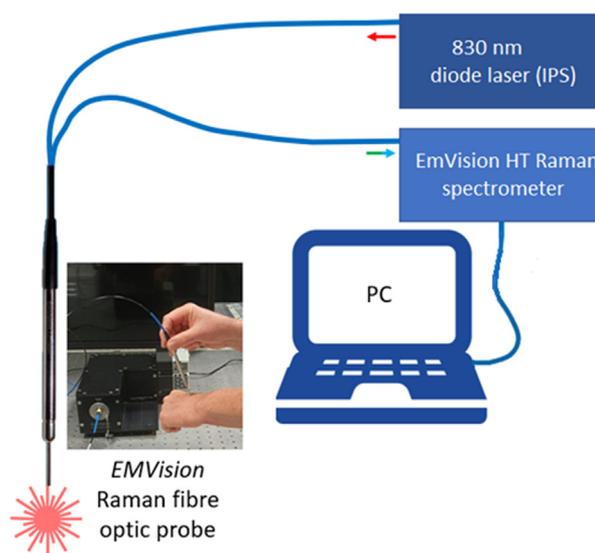
Figure 1. Schematic layout of Raman measurement system for in vivo skin recording of skin spectra (not drawn to scale) showing Raman probe and spectrometer. IPS, Innovative Photonic Solutions; PC, principal component.

dermatoses are subtle and due to different biological changes and biomarkers. These subtle intergroup spectral differences can easily be overwhelmed by intragroup spectral variations, which can degrade the stability and generalizability of a classifier. Being a supervised method, the partial least square (PLS) regression finds features and projections representing the intergroup differences in the new data space that best separate the different groups. PLS-DA can be prone to over-fitting; however, this can be checked by examining the spectral loadings.

The high ratio of benign versus malignant lesions in our dataset was a result of including as wide a variety of benign skin lesions and inflammatory dermatoses as possible. The aim was to include all possible samples that could be encountered in dermatological and general practitioner clinics, so as to provide a reliable, objective assessment of whether or not a lesion presented is malignant or benign, without having to rely only on clinical experience.

RESULTS AND DISCUSSION

Five different classification models were created from the Raman spectra of 867 samples that would provide a useful diagnostic aid for clinicians in a routine clinic. Four models were constructed from 4 training data subsets of the total set of Raman spectra and are listed below. For comparison, an additional classification model (V) was created for a particular subset of dataset I, which comprised only those lesions (benign as well as malignant) that had been biopsied with histology testing. Also investigated was the accuracy of classifying malignant melanomas from the 2 benign lesions more commonly misdiagnosed owing to their similar pigmented appearance. A training and validation subset I (ii) of dataset I (i) was therefore set up to include only the benign lesions seborrheic keratosis (SK) and different melanocytic nevi (congenital, compound, blue, intradermal, lentiginous junctional, halo, pedunculated and speckled lentiginous nevi) and malignant melanomas.

I. (i) all malignant lesions versus all benign lesions (ie, excluding inflammatory dermatoses) and (ii) malignant melanomas and selected melanocytic benign lesions: SK and different melanocytic nevi (congenital, compound, blue, intradermal, lentiginous junctional, halo, pedunculated, and speckled lentiginous nevi);
II. all malignant lesions versus all benign skin conditions (ie, including both benign lesions and inflammatory dermatoses);
III. all malignant lesions versus all inflammatory dermatoses;
IV. differentiation of melanomas from the NMSCs, BCC, and SCC;
V. all malignant lesions versus benign lesions with histology test results.

All malignant lesions versus different types of benign lesions
The mean spectra with SD for each of the malignant and benign lesions in data subset I (i) are shown in Figure 2a; the mean spectra had been normalized to the same height for the dominant 1446.5 cm^{-1} peak (−CH$_2$ deformation representing all lipid, keratin, and collagen components). The mean spectra have been overlaid in Figure 2b to enable a closer

comparison of the relative peak height differences, which represent the specific molecular composition in each group of lesion types. A difference plot of the 2 is given below the spectra, calculated by subtracting the average spectrum of the benign lesions from that of the malignant lesions; it has been magnified by a factor of 5 for a clearer view of the differences. The positive peaks indicate components present in greater amounts in the malignant lesions, whereas those in decreased amounts are represented by negative peaks, all relative to the CH$_2$ deformation mode (subject to the least change in intensity than other molecular functional groups). Negative peaks include those at 811, 858, 873, 919, 940, 973, 1167, 1146, 1182, and 1196 cm^{-1} (indicating reduction in the aromatic amino acids tyrosine and tryptophan as well as proline, hydroxyproline, and valine [Movasaghi et al, 2007; Stone et al, 2004]). Reduction in alpha-helix structures in the malignant lesions is also indicated by the negative peaks at 1241 and 1271 cm^{-1} (amide III) and 1500 and 1534 cm^{-1} (amide II vibrations). Similar changes have been observed in previous studies for melanoma and BCC using laboratory benchtop Raman instruments (Gniadecka et al, 2004). Also decreased in malignant lesions are fatty acids, shown by the negative peaks at 1383, 1401, and 1427 cm^{-1}. (Movasaghi et al, 2007; Stone et al, 2004).

The positive peak at 1003 cm^{-1} (representing phenylalanine) shows that these compounds are increased in malignant lesions. Also increased in malignant lesions are phospholipids (positive peaks at 957 and 1078 cm^{-1}) and DNA purine bases (1335 cm^{-1} [Huang et al, 2003; Movasaghi et al, 2007]). Increases in malignant lesions of lipids and ceramides are shown by the positive peaks at 1311 cm^{-1} (CH$_2$ and CH$_3$ twisting modes), 1125 cm^{-1} (C−C and C−N stretching), and 1650 and 1669 cm^{-1} (C=C and C=O) (Movasaghi et al, 2007).

Principal component (PC) analysis of the spectra did not show good separation between the classes because the differences between the spectra were very subtle. PLS-DA (with cross-validation) was therefore used to better differentiate and classify malignant versus benign lesions with 11 PCs used that explained 95% of the variance; the PLS loadings are shown in Figure 3.

The diagnostic accuracy of the classification model was evaluated in terms of the area under curve (AUC) of the receiver operator curve (ROC), as described in the Materials and Methods. Figure 2c shows the cross-validated ROC plots for the training set (532 benign, 96 malignant samples) with AUC and 95% confidence intervals (CIs); the AUC was 0.904. Figure 2d shows the associated bee-swarm plots for cross-validated prediction in the training set.

Figure 2e shows the ROC plots with AUC and 95% CIs for the separate validation set (82 benign, 26 malignant samples); AUC was 0.916. The corresponding predictions of the lesions in the validation set are shown as bee-swarm plots in Figure 2f.

The specificities at 100, 95, and 90% sensitivity levels are presented in Table 1 (left-hand side) for the training set and validation set, for each of data subsets I−V and include the number of PCs (latent variables) and the AUC for both the training sets and also the validation sets. Two different threshold levels were selected (between 1 and −1) for the

Figure 2. Differentiation between malignant and benign lesions, excluding inflammatory dermatoses. (**a**) Mean spectrum with SD for each of the benign (top) and malignant skin lesions. (**b**) Mean spectra of each malignant (purple) and benign (light blue) lesion, normalized to the same height for the −CH$_2$ deformation band at 1446.5 cm^{-1}. The red curve represents the difference between the malignant mean spectrum and the benign mean spectrum, amplified by a factor of 5 to enhance peak differences. The asterisk indicates an instrument peak. (**c**) Cross-validated ROC plots for training set, (**d**) bee-swarm plots for cross-validated prediction in training set (532 benign, 96 malignant samples), (**e**) ROC plots with 95% confidence intervals for validation set, and (**f**) bee-swarm plots for prediction of the separate validation set (82 benign, 26 malignant samples). arb., arbitrary; ROC, receiver operator curve.
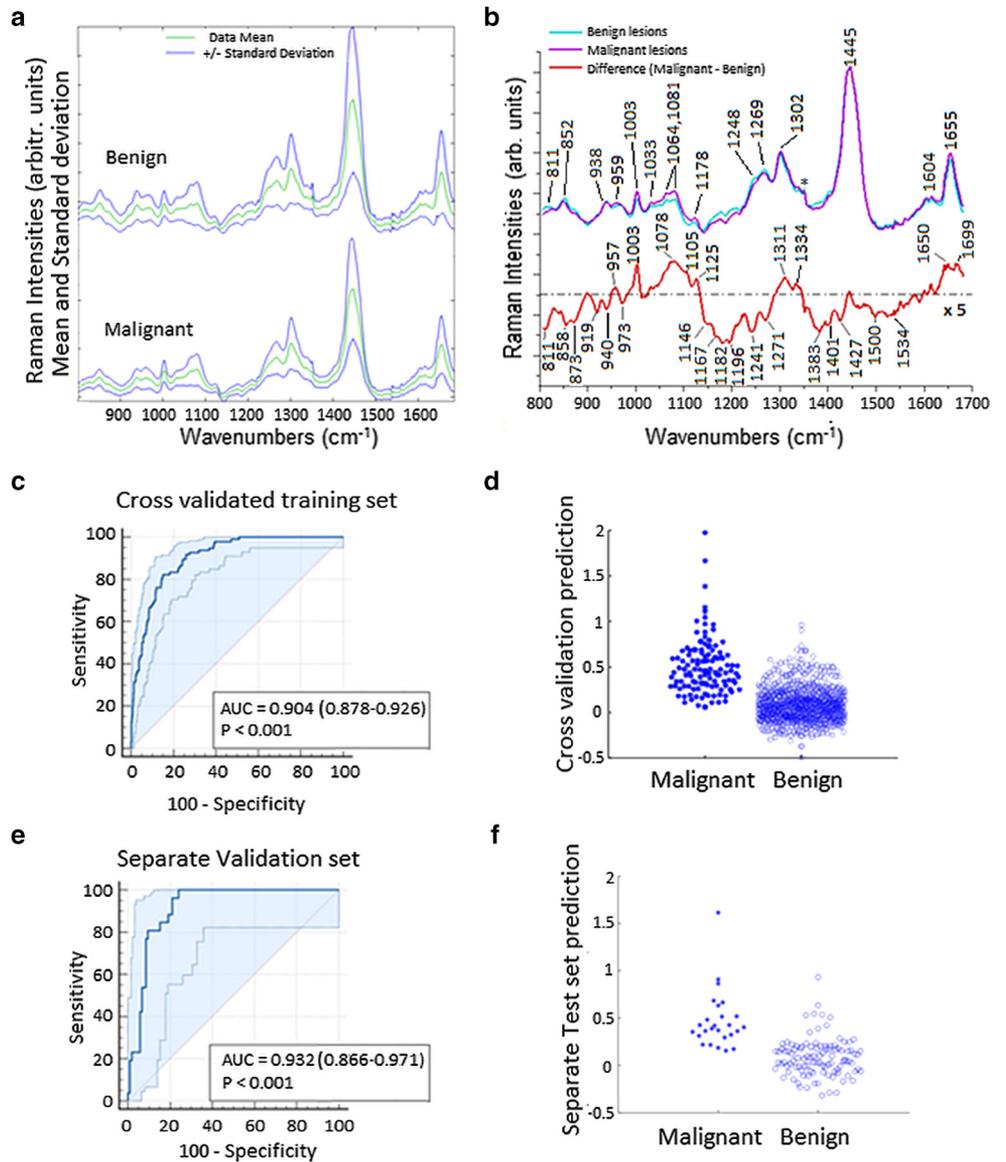
100% sensitivity levels for each training set to compare their predictive accuracy for the validation set. Threshold A was selected as the minimum predicted level that ensured 100% sensitivity of both the training set and the validation set, and threshold B was selected as the minimum predicted level that ensured 100% sensitivity of only the training set. The sensitivity and specificities for each threshold for each dataset are given on the right-hand side of Table 1. For this data subset I (i), the classification model showed good accuracy, achieving 100% sensitivity for prediction of malignant lesions versus benign lesions, with 51% specificity. This was achieved for threshold settings A and B.

For data subset I (ii), the mean spectra for the melanomas and the benign lesions with their SDs are shown in Figure 4a. The differences in peak shapes and relative peak heights between the 2 mean spectra in the figure as well as their SDs are subtle. The melanoma mean and SDs in Figure 4a show a slightly broader phenylalanine peak at 1003 cm^{-1} than the benign SK and nevi. The mean spectra are overlaid in Figure 4b, and the spectral differences between the 2 are

shown in the bottom curve. The difference spectrum appears similar in the 900 and 1300 cm^{-1} region to those seen in Figure 2b for all malignant lesions and benign lesions. The broadening of the phenylalanine 1003 cm^{-1} peak results in the appearance of the difference spectrum of 2 positive peaks at 993 and 1014 cm^{-1}. In addition, the positive peak at 1039 cm^{-1} in the difference spectrum is due to the C−H bending of phenylalanine; these suggest increased levels and changes in the molecular environment of phenylalanine in the malignant lesions. Positive peaks at 1058, 1078, 1095, and 1107 cm^{-1} (all lipid C−C stretching and symmetric PO$_2^-$ stretching vibration of phospholipids) show higher levels of nucleic acids and lipids in malignant melanoma than in the SK and nevi melanocytes.

The positive bands in the difference spectrum at 1305, 1434, 1507, 1522, 1540, and 1653 cm^{-1} represent increased levels of both protein and lipid contents in the malignant melanoma cells compared with those in the benign nevi and SK lesions (Movasaghi et al, 2007). The positive peaks between 993 and 1100 cm^{-1} and the amide I peak at 1653 cm^{-1} for the
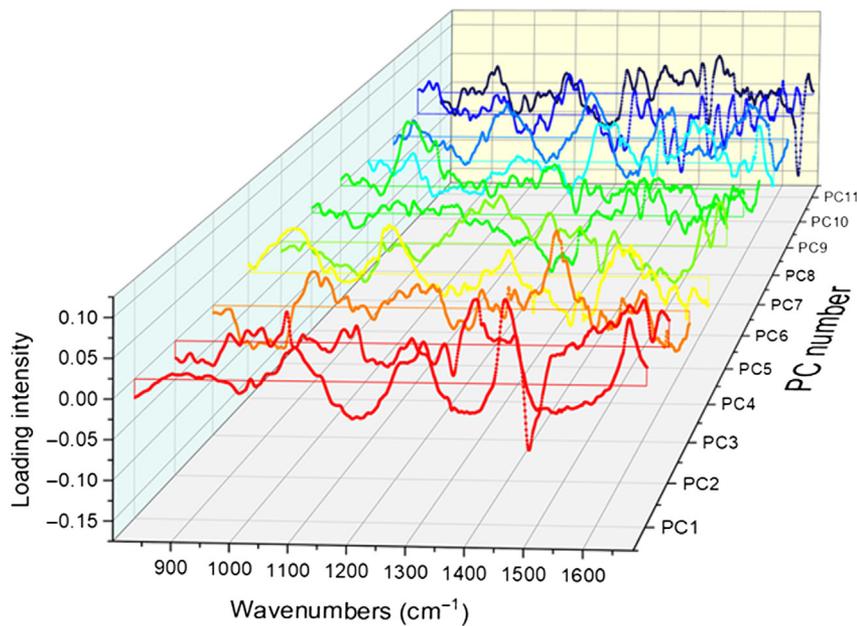
malignant melanoma agree with major peak differences found in a Raman microscopy investigation of benign melanocytes and different malignant melanoma cell lines (SKMEL28, MALME-3M, SKMEL5, UACC-62, and 451LU) (Brauchle et al, 2014), where the melanocytes could be distinguished from melanoma cells from differences in the peaks for phenylalanine, amide I, and DNA, including phospholipids.

A cross-validated PLS-DA classification model was constructed using 8 PCs (explaining 93% of the variance) for differentiating melanomas from these 2 benign classes from the data subset I (ii) of spectra of melanomas, SK, and the nevi. Figure 4c shows the cross-validated ROC plots for the training set (352 benign, 23 malignant samples) with an AUC of 0.845; CIs are indicated on the plot and also given in Table 1. The cross-validated predictions by the training model for classifying lesions as melanomas versus SK and nevi are given as bee-swarm plots in Figure 4d. The ROC for the separate validation set (126 benign, 4 malignant samples) with 95% CIs is given in Figure 4e with an AUC of 0.956. The predictions of lesions in the separate validation set as melanomas versus SK and nevi are given as bee-swarm plots in Figure 4f. The 8 PLS loadings are shown in Figure 5.

The AUC obtained for the validation set and specificities of 90% at both 90 and 95% sensitivities (Table 1) are better than those obtained for dataset I (i) shown in Figure 2e. However, the specificity of 6% is much lower for the 100% sensitivity level and also lower than the 15% obtained at 99% sensitivity in a previous study (Lui et al, 2012) for a similar dataset. This is due to a particular malignant sample in the PLS-DA bee-swarm plot in Figure 4f showing a lower predicted value for malignancy than the other 3, which unlike the other melanoma samples, does not show the characteristic brown pigmentation of melanin within the erythematous papule and so is described as amelanotic (lesion 2 in Figure 6; lesion 1 was confirmed as a compound nevus). Inclusion of more such amelanotic melanomas in this dataset of melanomas, SK, and nevi would improve the classification accuracy. Notably, this lesion was predicted more accurately in dataset I (i), which included other malignant and benign amelanotic lesions.

Malignant lesions versus all benign samples (skin lesions and inflammatory dermatoses)

This is the most useful triage for a primary clinician, who needs to decide whether or not biopsy/excision is indicated. The PLS-DA classification of malignant lesions from both benign skin lesions and inflammatory dermatoses is represented for the training and validation sets in Figure 7a and b and c and d, respectively. Twelve PCs were used for the model, explaining 95% of the variance; the PLS loadings are shown in Figure 8. Good accuracy was obtained for the cross-validated training classification model, with an AUC of 0.904 (95% CI = 0.880−0.924) (Figure 7a) with a specificity of 50% at 100% sensitivity. Figure 7b shows bee-swarm plots for the cross-validated prediction model results (641 benign [both inflammatory dermatoses and benign lesions, 96 malignant lesions] with an AUC of 0.904 [95% CI = 0.880−0.924]). The bee-swarm plots represent the predicted values for the 96 malignant lesions and 641 benign skin samples (both lesions and inflammatory dermatoses).

The ROC plots in Figure 7c with 95% CIs and associated bee-swarm plots in Figure 7d obtained for prediction of the separate validation set of 26 malignant skin lesions and 104 benign skin samples show high prediction accuracy, with an AUC of 0.916 (95% CI = 0.855−0.958) (Figure 7c) and 100% sensitivity and 50% specificity (Table 1, right-side hand). The specificities of 50, 50, and 70% obtained respectively for 100, 95, and 90% sensitivities were comparable with those for the skin lesions only (Table 1), which are better than those obtained in another study on a similar dataset (Lui et al, 2012).

6

JID Innovations (2024), Volume 4

M Nieuwoudt et al.
Raman Device Differentiates Cancer and Benign Skin Lesions

Table 1. PLS-DA Classification Specificities at Sensitivities of 100, 95, and 90% and ROC AUCs for Training and Validation Data Subsets I−V

Data set	Number of PCs (% variable)	Specificity at Threshold A Set at Level to Ensure 100% Sensitivity for Both CV Training and Validation Sets		Specificity at Threshold B Set at Level to Ensure 100% Sensitivity for CV[6] Training Set Only		Specificity at 95% Sensitivity		Specificity at 90% Sensitivity		AUC	AUC	Accuracy for Prediction of Validation Set Using Threshold A		Accuracy for Prediction of Validation Set Using Threshold B	
		Training CV (95% CI)	Validation (95% CI)	Training CV[6] (95% CI)	Validation (95% CI)	Training CV (95% CI)	Validation (95% CI)	Training CV[6] (95% CI)	Validation (95% CI)	Training CV (95% CI)	Validation (95% CI)	Sensitivity	Specificity	Sensitivity	Specificity
I (i)[1]	11 (95%)	47% (39−52)	51% (43−61)	47% (39−52)	51%(43−61)	58% (46−74)	85% (73−93)	75% (59−86)	86% (72−93)	0.904 (0.878−0.926)	0.932 (0.866−0.971)	100% (86−100)	51% (43−61)	100% (86−100)	51% (43−61)
I (ii)[2]	8 (93%)	28% (6−48)	6% (2−26)	28% (6−48)	6% (2−26)	48% (8−64%)	90% (80−96)	59% (10−76)	90% (81−96)	0.845 (0.805−0.880)	0.956 (0.906−0.984)	100% (40−100)	6% (2−26)	100% (40−100)	6% (2−26)
II[3]	12 (95%)	43% (40−54)	50% (35−63)	43% (40−54)	50% (35−63)	61% (46−71)	70% (69−89)	74% (58−79)	79% (67−87)	0.904 (0.880−0.924)	0.916 (0.855−0.958)	100% (87−100)	50% (35−63)	100% (87−100)	50% (35−63)
III[4]	13 (95%)	36% 26-47	47% (20−73)	36% (26−47)	47% (20−73)	61% (51−80)	73% (33−97)	80% (56−88)	80% (33−97)	0.926 0.878−0.960	0.929 0.825−0.981	100% (86−100)	47% (20−73)	100% (86−100)	47% (20−73)
IV[5]	11 (98%)	54% (37−80)	50% (36−70)	54% (37−80)	50%(36−70)	54% (38−80)	87% (56−100)	92% (74−100)	87% (56−100)	0.945 (0.880−0.981)	0.948 (0.762−1)	100% (54−100)	50% (36−70)	100% (54−100)	50% (36−70)
V[6]	15 (96%)	40% (22−58)	50% (2−65)	93% (86−99)	36% (13−65)	69% (55−78)	50% (14−86)	79% (67−88)	93% (43−100)	0.921 (0.876−0.953)	0.896 (0.717−0.978)	100% (75−100)	50% (2−65)	77% (46−95)	93% (43−100)

Abbreviations: AUC, area under curve; BCC, basal cell carcinoma; CI, confidence interval; CV, cross-validation; PC, principal component; PLS-DA, partial least squares discriminant analysis; ROC, receiver operator curve; SCC, squamous cell carcinoma; SK, seborrheic keratosis.

The % variable denotes the percentage variance explained by these PCs.

[1]Malignant versus all benign skin lesions.

[2]Melanomas versus SK and nevi only.

[3]Malignant lesions versus all benign skin conditions.

[4]Malignant lesions versus all inflammatory dermatoses.

[5]Melanoma versus BCC and SCC.

[6]Malignant versus histology-confirmed only benign lesions.
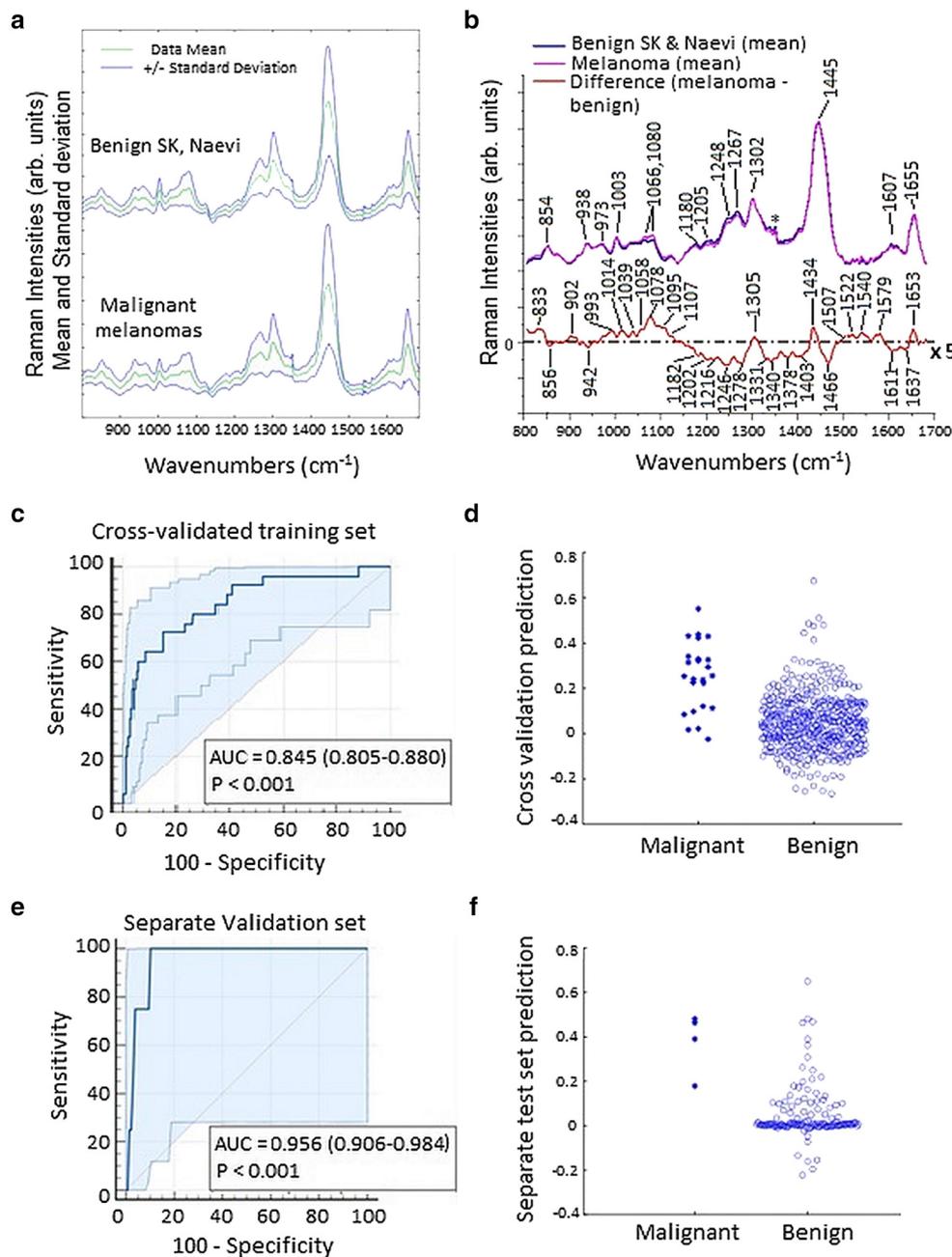
Figure 4. Spectral differences between malignant melanomas and different types of benign nevi and SK.
(**a**) Mean spectrum with SD for each of the benign (top) and malignant melanoma skin lesions. (**b**) Mean spectra of each malignant (purple) and benign (light blue) lesion, normalized to the same height for the $-CH_2$ deformation band at 1446.5 cm^{-1}. The red curve represents the difference between the malignant mean spectrum and the benign mean spectrum, amplified by a factor of 5 to enhance peak differences. The asterisk indicates a very sharp peak that is unassigned and presumably due to a faulty pixel on the CCD detector. (**c**) Cross-validated ROC plots for training set, (**d**) bee-swarm plots for cross-validated prediction in training set (352 benign, 23 malignant samples), (**e**) ROC plots with 95% confidence intervals for validation set, and (**f**) bee-swarm plots for prediction of the separate validation set (126 benign, 4 malignant samples). arb., arbitrary; CCD, charged couple device; ROC, receiver operator curve; SK, seborrheic keratosis.

Malignant lesions versus inflammatory dermatoses

For this model, 13 PCs were used that explained 96% of the variance; the PLS loadings are shown in Figure 9. Good predictive ability for identifying malignant lesions from the inflammatory dermatoses was obtained for the cross-validated training set as seen by the AUC of 0.926 (95% CI = 0.878−0.960) for ROC and corresponding bee-swarm plots in Figure 10a and b, respectively. The training set comprised 96 malignant lesions and 86 benign inflammatory dermatoses.

The ROC plot of the separate validation set (26 malignant skin lesions and 32 benign inflammatory dermatoses) is given in Figure 10c; the AUC for prediction of the validation set was 0.929 (95% CI = 0.825−0.981). The corresponding bee-

swarm plots for prediction of the validation set are given in Figure 10d.

The validation set was predicted with very high accuracy: 100% sensitivity and 47% specificity (Table 1) and a biopsy ratio of 1.6. This classification would be particularly useful for distinguishing those benign inflammatory dermatoses that appear similar to malignant lesions. For example, the differential diagnosis of a red patch on chronically sun-exposed skin is wide, including carcinoma in situ; superficial BCC; and inflammatory dermatoses such as eczema, psoriasis, tinea, and less commonly discoid lupus erythematosus. Support establishing the correct diagnosis by Raman spectroscopy will help those without specialist training in cutaneous medicine, especially community physicians. To our

Figure 5. PLS loadings for the 8 PCs used in the classification training set model for differentiating melanomas from benign SKs and nevi, which explained 93% of the total data variance. PC, principal component; PLS, partial least square; SK, seborrheic keratosis.
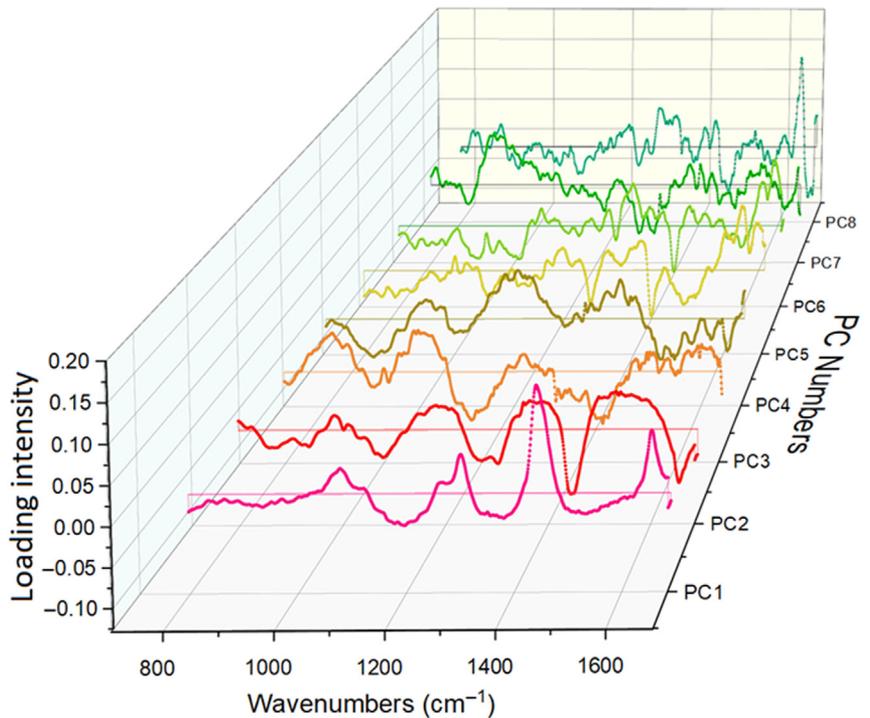
knowledge, the Raman differentiation in this study of a large variety of inflammatory dermatoses from malignant lesions has not been reported previously.

Melanomas versus NMSCs (BCC and SCC)
This classification was used to differentiate melanomas from NMSCs. The classification is needed so that melanoma management can be prioritized above others, especially in a resource-constrained environment. Amelanotic melanomas can be difficult to diagnose even for experienced practitioners such as that shown in Figure 4.

The mean spectrum with SD for each of the melanomas and both NMSCs are shown in Figure 11a; the mean spectra had been normalized to the same height for the dominant 1446.5 cm^{-1} peak ($-CH_2$ deformation representing all lipid, keratin, and collagen components). The mean spectra of

Figure 6. Photographic images of (1) compound nevus and (2) amelanotic melanoma.

melanoma and NMSCs (BCC and SCC) are shown overlaid in Figure 11b to enable a closer comparison of the relative peak height and shape differences that represent the specific molecular composition in each group of lesion types. The red plot below is the difference plot magnified by a factor of 5 and shows more clearly which components occur more or less in each.

Melanoma shows relatively lower peak intensities at 961 cm^{-1} (due to lipids), 1003 cm^{-1} (phenylalanine), and 1656 cm^{-1} (amide I band) than BCC and SCC. Protein structures in malignant cells are reduced compared with those in normal cells (Gniadecka et al, 2004; Huang et al, 2001); the lower intensities for protein peaks observed in Figure 11a for melanomas than for the NMSCs suggest that changes in the protein structure occur to a greater extent in melanoma than in NMSC lesions. However, increased peak intensities for the melanoma mean spectrum compared with those for the NMSCs appear at 820, 849, and 1176 cm^{-1} (tyrosine) and 1199 and 1210 cm^{-1} (hydroxyproline and tyrosine) (Brauchle et al, 2014; Movasaghi et al, 2007; Puppels et al, 1991).

Phenylalanine and tyrosine are free aromatic amino acids known to be involved in melanogenesis (Schallreuter and Wood, 1999); the observation of reduced phenylalanine but increased tyrosine and hydroxyproline appears to be characteristic of melanoma. The observation of increased tyrosine peaks in melanoma in this study is consistent with a report of increased levels of phosphotyrosine in melanoma, due to elevated protein tyrosine kinase activity (McArdle et al, 2001). The increased intensities over the regions in Figure 11b between 1380−1403 cm^{-1} and 1500−1550 cm^{-1} are likely due to higher levels of melanin in the melanocytes, which is characterized by broad bands centered around 1350 and 1580 cm^{-1}. The reduced intensities in
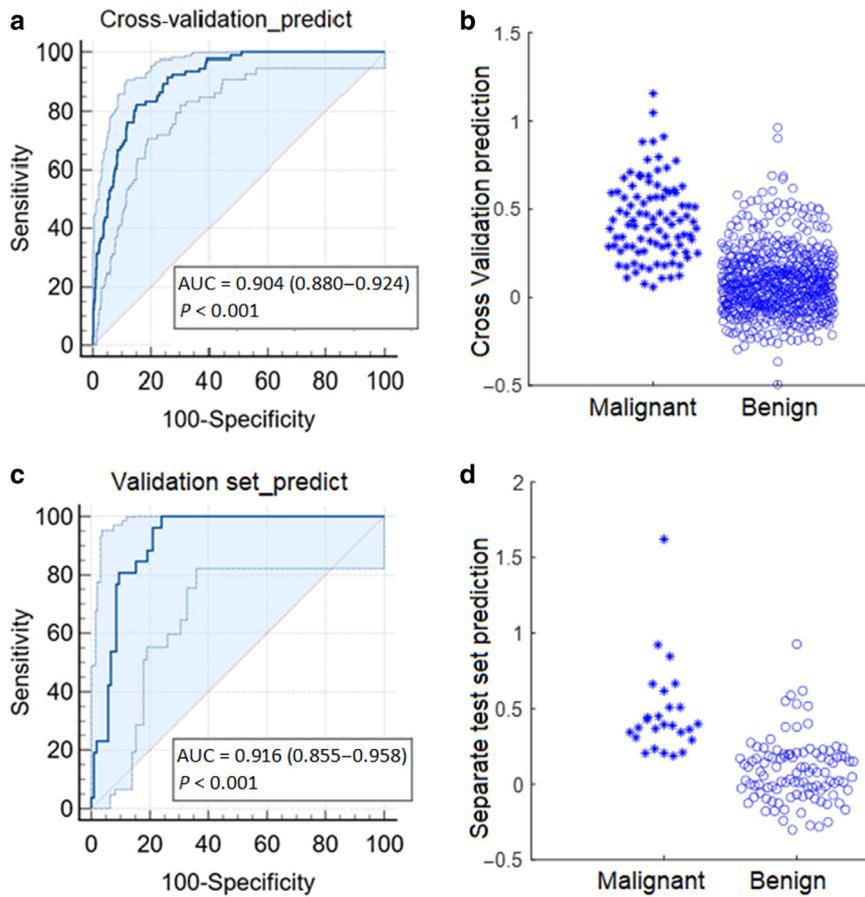
Figure 7. PLS-DA differentiation between malignant skin lesions and all benign skin samples (both lesions and inflammatory dermatoses). (**a**) ROC plots with 95% CIs and (**b**) bee-swarm plots for cross-validated prediction model results (641 benign, 96 malignant) with AUC of 0.904 (95% CI = 0.880−0.924). (**c**) ROC plots with 95% CIs and (**d**) associated bee-swarm plots for prediction of the separate validation set (104 benign skin samples and 26 malignant lesions) with AUC of 0.916 (95% CI = 0.855−0.958). AUC, area under the curve; CI, confidence interval; PLS-DA, partial least squares discriminant analysis; ROC, receiver operator curve.

melanomas shown by the negative peaks in the difference spectrum at 1003 and 1103 cm^{-1} indicate that the NMSCs have higher levels of phenylalanine than melanomas.

The cross-validated ROCs for the differentiation of melanomas from NMSCs are given in Figure 11c for the training set (77 NMSCs, 21 melanomas) with an AUC of
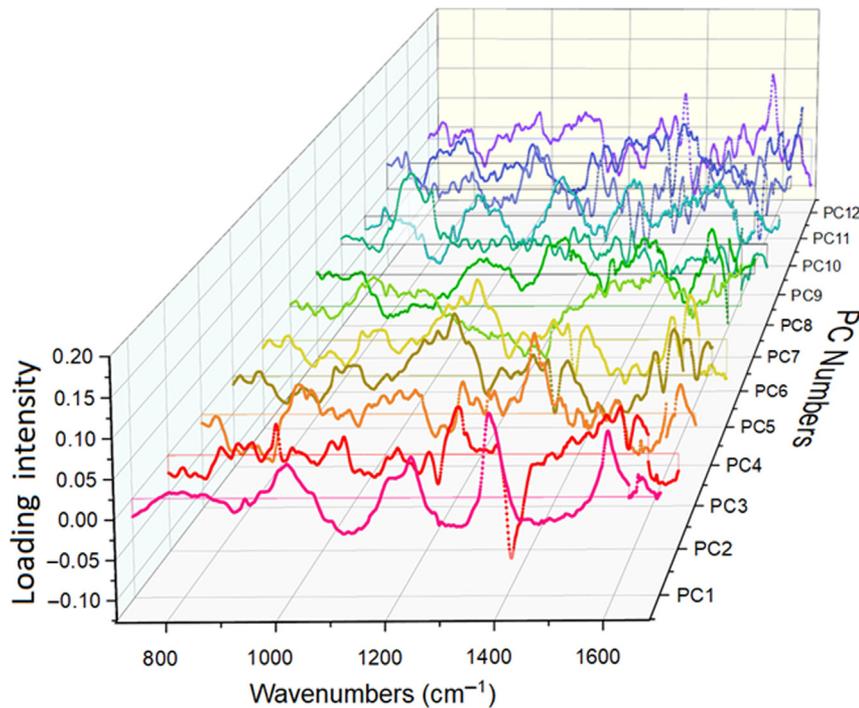
Figure 8. PLS loadings for the 12 PCs used in the classification training set model for differentiating all malignant lesions from both benign lesions and inflammatory dermatoses, which explained 95% of the total data variance. PC, principal component; PLS, partial least square.

Figure 9. PLS loadings for the 13 PCs used in the classification training set model for differentiating malignant lesions from inflammatory dermatoses, which explained 96% of the total data variance. PC, principal component; PLS, partial least square.
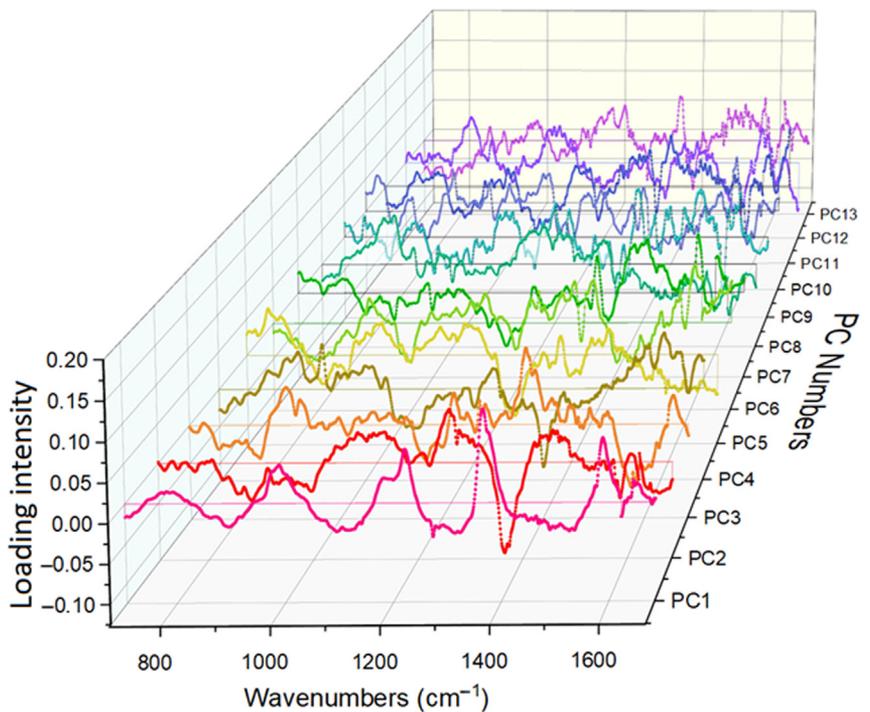
0.945 (95% CI = 0.880−0.981). The associated bee-swarm plots are given in Figure 11d. The ROC plot for the separate validation set (16 NMSCs, 6 melanomas) with an AUC of 0.948 (95% CI = 0.762−1) is given in Figure 11e, and the associated bee-swarm plots are given in Figure 11f. The amelanotic sample that was poorly predicted in Figure 4f is better differentiated as malignant in Figure 11f using the classification model from this

Figure 10. Differentiation between malignant lesions and benign inflammatory samples. (**a**) ROC plots with 95% CIs and (**b**) bee-swarm plots for cross-validated prediction of the training set comprising 96 malignant lesions and 86 benign inflammatory dermatoses, with AUC of 0.926 (95% CI = 0.878−0.960). (**c**) ROC plot with 95% CIs for prediction of the validation set of 26 malignant skin lesions from 32 benign inflammatory dermatoses with AUC of 0.926 (95% CI = 0.878−0.960). (**d**) Corresponding bee-swarm plots showing prediction results of the validation set. AUC, area under the curve; CI, confidence interval; ROC, receiver operator curve.
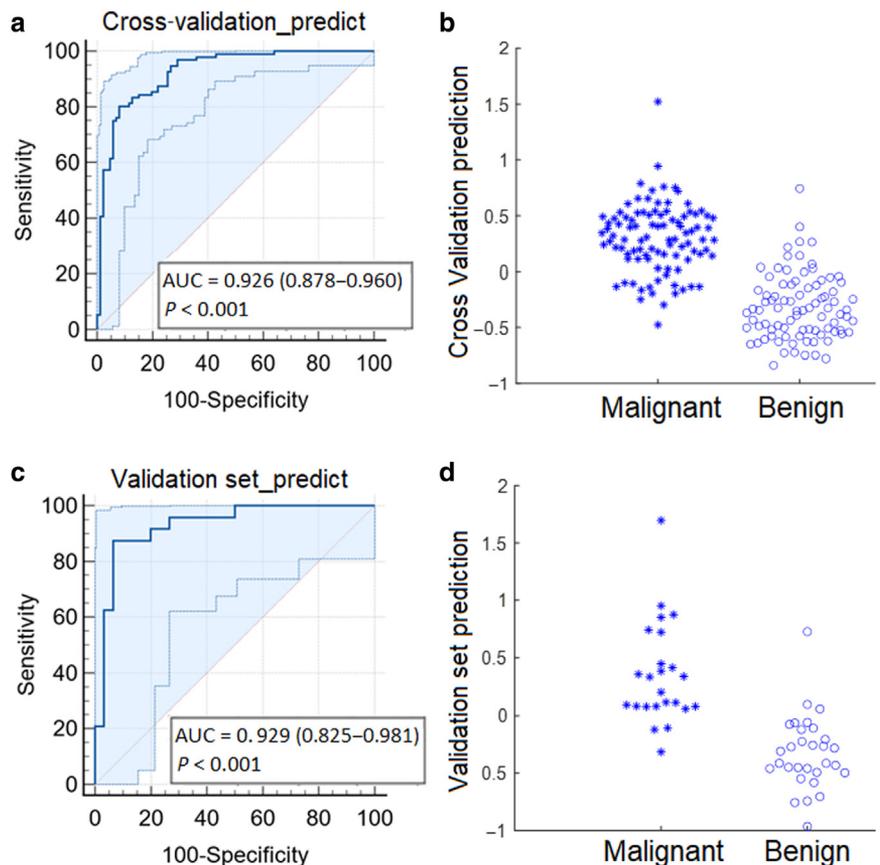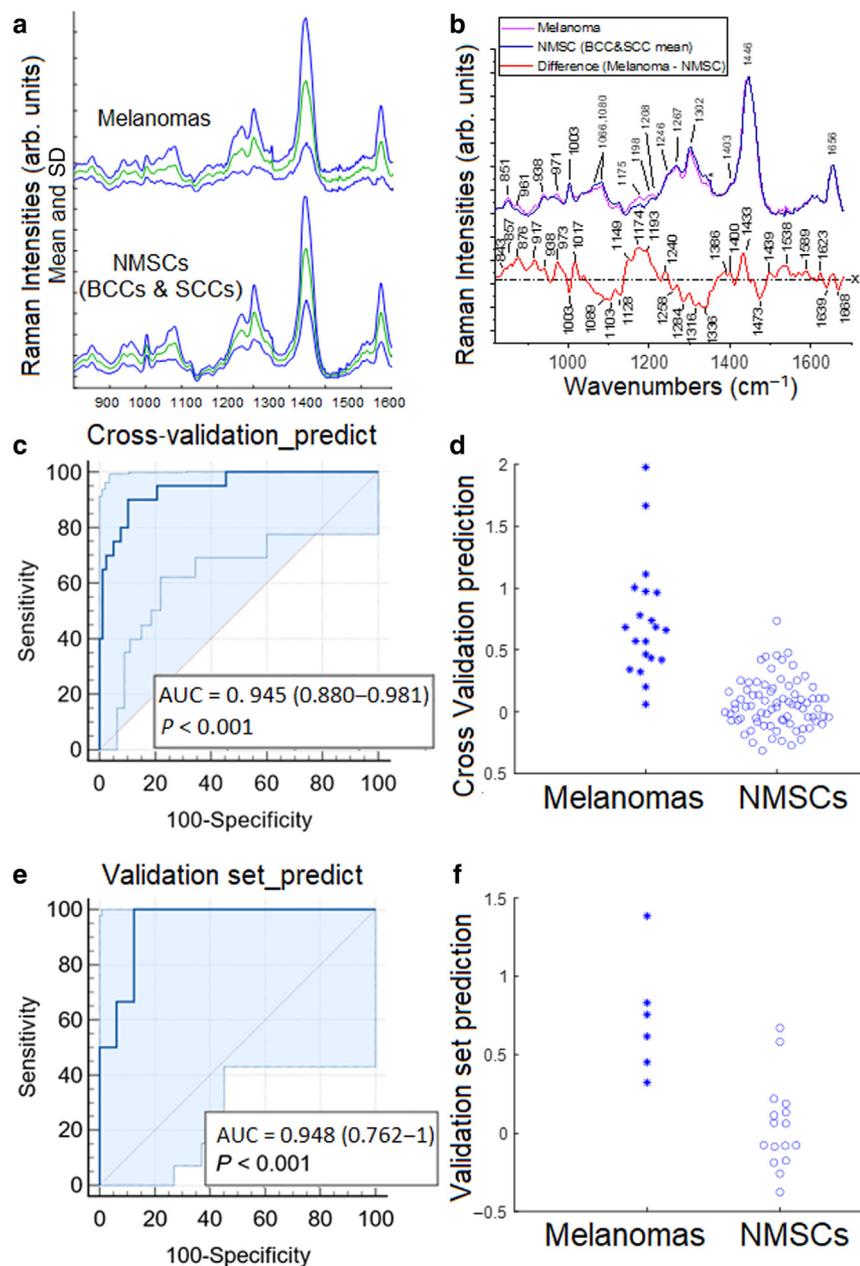
Figure 11. Differentiation between melanomas and NMSCs (BCCs and SCCs). (**a**) Mean spectrum with SD for each of all melanomas (top) and NMSCs (BCCs and SCCs) (bottom). (**b**) Mean spectra of the melanoma (pink) and NMSCs (purple) samples, normalized to the same height for the $-CH_2$ deformation band at 1446.5 cm^{-1}. The red curve represents the difference between the melanomas and the NMSCs amplified by a factor of 5 to enhance peak differences. The asterisk indicates an instrument peak. (**c**) Cross-validated ROC plots for training set with AUC of 0.945 (95% CI = 0.880−0.981), (**d**) associated bee-swarm plots for cross-validated prediction in training set of 77 NMSCs and 21 melanomas, (**e**) ROC plots for validation set with AUC of 0.948 (95% CI = 0.762−1), and (**f**) associated bee-swarm plots for prediction of the separate validation set (16 NMSCs, 6 melanomas). AUC, area under the curve; BCC, basal cell carcinoma; CI, confidence interval; NMSC, nonmelanoma skin cancer; ROC, receiver operator curve; SCC, squamous cell carcinoma;

particular dataset. The AUC values obtained for both training and validation sets were better than those in other studies (Khristoforova et al, 2019; Lui et al, 2012). The 12 PCs used for the PLS-DA model are given in Figure 12 and explain 95% of the variance.

Malignant and histologically tested-only benign lesions

This classification model was created for a particular subset of dataset I (i), for which only those lesions that had been biopsied (benign and malignant) with histology testing were included. These lesions were suspected as malignant and referred by the patients' general practitioners for biopsy, totaling 247. Before biopsy, they were further triaged by a plastic surgeon at the clinic. Those selected for biopsy underwent histology testing; 124 of these were confirmed as malignant, and 123 were confirmed as benign, indicating a clinical assessment biopsy ratio of 0.5. The training dataset comprised 220 of the samples—111 malignant and 109 benign—and the validation set had 27 samples (13 malignant and 14 benign).

For this PLS-DA model, 16 PCs were used that explained 96% of the variance; the PLS loadings are shown in Figure 13.

The ROC for the cross-validated training set (109 benign and 111 malignant lesions) is shown in Figure 14a for the training set with an AUC of 0.921 (95% CI = 0.876−0.953). Figure 14b gives the associated bee-swarm plots for the cross-validated training set.

Figure 12. PLS loadings for the 11 PCs used in the classification training set model for differentiating melanomas from nonmelanoma skin cancers; the 11 PCs explained 95% of the total data variance. PC, principal component; PLS, partial least square.
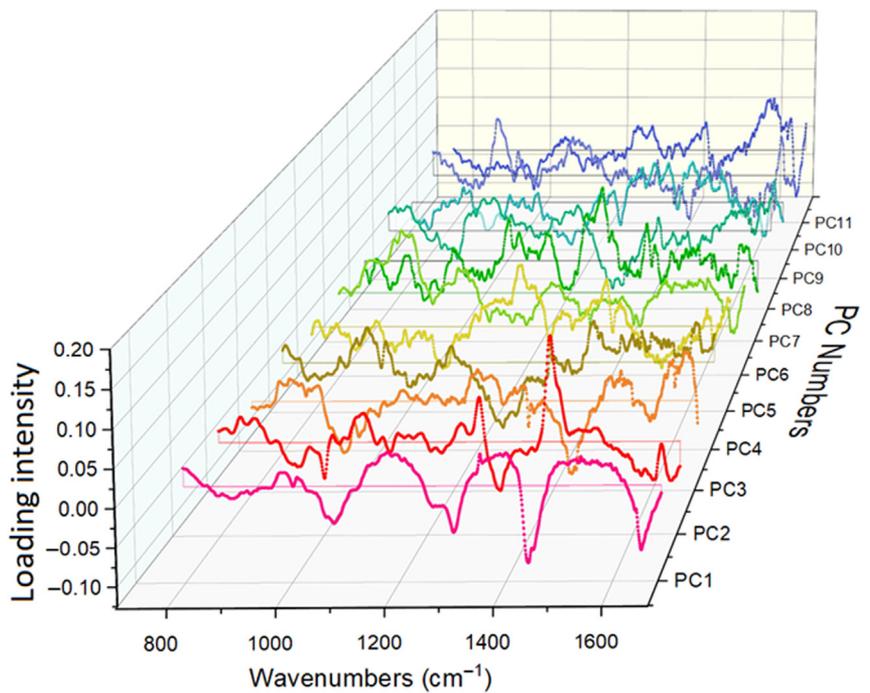
Figure 14c shows the ROC plot for the separate validation set (13 benign and 14 malignant lesions) with an AUC of 0.896 (95% CI = 0.880−0.981); Figure 14d shows the associated bee-swarm plots for prediction of this set. The high AUC of 0.921 with 93% specificity at 100% sensitivity is much higher than for the data subsets I−IV. However, this model predicted the validation set with 77% sensitivity and 93% specificity. In a previous study that compared the accuracy of prediction for a dataset comprising histology-only confirmed lesions with a dataset that also included clinician-diagnosed benign lesions (Lui et al, 2012), similar accuracy was obtained for the cross-validated classification.

However, for the training data subset V, the specificity at 100% sensitivity was 93%, and the prediction threshold set was 0.44 (threshold B in Table 1), which is high above the average of −0.06 for the other training subsets I−IV. The 100% threshold value needed for the 100% sensitivity prediction of the subset V validation set was 0.04; setting the threshold to this value (threshold A in Table 1) enabled 100% sensitivity for the validation set with 50% specificity

Figure 13. PLS loadings for the 16 PCs used in the classification training set model for differentiating the malignant lesions from the histology-only benign lesions; the 16 PCs explained 96% of the total data variance. PC, principal component; PLS, partial least square.

(Table 1). For the other datasets I−IV, the average threshold for the validation sets is 0.12. This highlights the need to include as many samples as possible in the training datasets to ensure accurate prediction of the validation sets.

Comparison of the effect of the different ethnic groups on the classification accuracy is limited by the different number of samples in each, with Europeans dominating the numbers at 70% and Latin American, African, and Middle Eastern groups having the lowest numbers at 4%. However, the cross-validated training set specificity at 100% sensitivity is comparable with those for the European (45%) and Asian (44%) groups; lowest for the Maori and Pacific Island group (37%); and highest for the Latin American, African, and Middle Eastern groups (73%) (Table 2). However, for the Latin American, African, and Middle Eastern groups validation set, no malignant lesions were present. For the other 3 groups, the best specificity at 100% sensitivity was achieved for the Asian group at 50%, followed by the European group at 44% and Maori and Pacific Island group at 33%.

The results in this study show high sensitivities and specificities for classification of malignant skin lesions from both benign skin lesions and inflammatory dermatoses as well as malignant melanomas from NMSCs. Comparison of the results with the accuracy of clinical diagnosis for malignant versus benign lesions and/or inflammatory dermatoses of other studies in the literature is hampered by the wide range and different measures of accuracy and different lesion types included in the sample sets used as well as the different levels of experience of clinicians and clinical practices reported (Heal et al, 2008; MacKenzie-Wood et al, 1998; Morton and Mackie, 1998; Trevethan, 2017; Ying et al, 2020). The more

useful measure in this study of the specificity obtained at 100% sensitivity for clinicians is given in Table 1 for each of the 5 data subsets as well as for the sensitivity settings at 95 and 90%. Using the 100% sensitivity thresholds for each of the 5 training data subsets, the sensitivities, specificities, and AUCs were determined for predicting the associated separate validation sets and are summarized in Table 1. The numbers of true- and false-positive and -negative samples are given in Table 3. Compared with the specificities obtained at 90, 95, and 99% sensitivities reported in a similar study by Lui et al (2012), the results summarized in this paper in Table 1 show improved specificities at the 90, 95, and 100% sensitivities for the calibration training sets and the separate validation sets.

The ROCs obtained from the 5 data subsets in this study show improved AUC values compared with those from a recent in vivo study using a different portable system (Bratchenko et al, 2021), in which AUCs of 0.75 (95% CI = 0.71−0.79) were reported for classification of malignant versus benign tumors, 0.69 (95% CI = 0.63−0.76) for melanomas versus pigmented tumors, and 0.81 (95% CI = 0.74− 0.87) for melanomas versus SK, respectively.

This research has developed a portable Raman system that will assist the clinician by providing an improved diagnostic tool to differentiate between benign and malignant tissues. This will translate and expand the scope of Raman spectroscopy with its inherent sensitivity and quantitative capacity beyond a laboratory and only experienced technical users. Through the application of the multivariate statistical classification algorithm, PLS-DA, on the spectra obtained with the

Table 2. Summary of the Specificity Obtained at 100% Sensitivity for the Cross-Validated Training Set for the Different Ethnicity and Lesion Types in Dataset II

	Training Set			Validation Set		
Ethnicity	**Number of Samples**	**Sensitivity**	**Specificity**	**Number of Samples**	**Sensitivity**	**Specificity**
European	499	100%	45.3%	90	100%	44.3%
Māori/Pacific peoples (Cook Island Māori, Samoan, Tongan)	74	100%	37.3%	21	100%	33.3%
Asian (Indian, Sri Lankan, Filipino, Chinese, Thai, Malaysian)	130	100%	43.9%	18	100%	50.0%
Latin American, African, and Middle Eastern	33	100%	72.7%	2	N/A	N/A

Abbreviation: N/A, not applicable.

portable system, it was possible to noninvasively differentiate the malignant lesions from benign lesions as well as inflammatory dermatoses. These differences could be explained by changes in their protein, lipid, and aromatic amino acid abundances that were evident from the Raman spectral differences between the mean malignant and benign samples and which agreed with those found in other studies in the literature.

The most useful diagnostic aid for a primary clinician is one that can help decide whether or not to biopsy when presented with a lesion that could be malignant, benign, or an inflammatory dermatosis. In this study, 100% sensitivity was achieved for predicting malignant lesions from benign and inflammatory dermatoses, with 50% specificity. The high accuracy obtained demonstrates the suitability of our portable Raman system for deployment in dermatological and general practitioner clinics to provide a reliable,

objective assessment of whether or not a lesion or inflammatory dermatosis presented is malignant or benign.

MATERIALS AND METHODS
Raman spectra
All Raman spectra were recorded in vivo using a portable Raman spectrometer (HT Raman spectrometer, EmVision LLC) with a custom-made fiber optic probe. A schematic diagram of the Raman system layout is shown in Figure 1. The probe was placed on the lesion of interest, and then spectra were recorded using the Andor Solis software and processed using MATLAB and PLS-Toolbox 9.1 (Eigenvector Research). The excitation radiation used was an 830 nm central wavelength solid-state diode laser (Innovative Photonic Solutions) with 40 mW output power launched into the probe; the power density measured at the skin was 1.06 J/cm^2, which is below the maximum permitted exposure of 1.67 J/cm^2 as specified by the American National Standards Institute (Choquette et al, 2007). Each

Table 3. Summary of Confusion Table Data for 100% Sensitivity for Each Data Subset

Training Set	100% Sensitivity		Specificity	PPV at 100% Sensitivity	Validation Set	100% Sensitivity		Specificity
Training Set					**Test Set**			
Dataset I (i)	TP	96	46.8%	25%	Dataset I (i)	TP	26	51.2%
	FP	283				FP	40	
	TN	249				TN	42	
	FN	0				FN	0	
Dataset I (ii)	TP	23	27.8%	8%	Dataset I (ii)	TP	4	5.6%
	FP	254				FP	119	
	TN	98				TN	7	
	FN	0				FN	0	
Dataset II	TP	96	43.2%	21%	Dataset II	TP	26	50.0%
	FP	358				FP	53	
	TN	272				TN	51	
	FN	0				FN	0	
Dataset III	TP	96	36%	64%	Dataset III	TP	26	46.7%
	FP	55				FP	16	
	TN	31				TN	14	
	FN	0				FN	0	
Dataset IV	TP	20	54.5%	36%	Dataset IV	TP	7	50.0%
	FP	35				FP	8	
	TN	42				TN	8	
	FN	0				FN	0	
Dataset V	TP	111	40.4%	63%	Dataset V	TP	13	35.7%
	FP	65				FP	9	
	TN	44				TN	5	
	FN	0				FN	0	

Abbreviations: FN, false negative; FP, false positive; PPV, positive predictive value; TN, true negative; TP, true positive.

Table 4. List of Different Types of Lesions and Inflammatory Dermatoses Measured and Included in the Datasets

Group	Type	Individual Type	Number Measured
Melanocytic benign lesions	Nevus	Blue nevus	2
		Café au Lait	8
		Compound nevus	6
		Congenital nevus	2
		Ephelides	14
		Halo nevus	2
		Intradermal nevus	7
		Junctional nevus	5
		Lentiginous nevus	18
		Lentigo	18
		Nevus otherwise not specified	286
		Pedunculated nevus	1
		Solar lentigo	3
		Speckled lentiginous nevus	1
Nonmelanocytic benign lesions	Other	Accessory nipple	2
		Acrochordon	1
		Actinic keratosis	41
		Angiokeratoma	1
		Angioma	1
		Becker nevus	3
		Cherry angioma	1
		Dermatofibroma	9
		Haemangioma	37
		Keratoacanthoma	4
		Keratosis pilaris	1
		Linear epidermal nevus	2
		Milia cyst	1
		Myxoid cyst	1
		Pilomixatrona	1
		Plantar fibromatosis	1
		Scar tissue	11
		Scar with no residual malignancy	16
		Sebaceous hyperplasia	9
		Seborrheic keratosis	86
		Vascular malformation	1
Malignant lesions	Melanoma	Acral melanoma	1
		Lentiginous melanoma	2
		Melanoma in situ	12
		Metastatic melanoma	3
		Spitzoid melanoma	1
		Superficial spreading melanoma	8
	SCC	SCC well differentiated	3
		SCC in situ	7
		SCC poorly differentiated	2
		SCC moderately differentiated	11
		SCC otherwise not specified	8
	BCC	BCC in situ	2
		BCC nodular	32
		Superficial BCC	14
		Infiltrative BCC	5
		Multifocal BCC	2
		BCC otherwise not specified	11

(*continued*)

Table 4. Continued

Group	Type	Individual Type	Number Measured
Inflammatory dermatoses	Eczema	Atopic eczema	17
		Discoid eczema	3
		Dermatitis not otherwise specified	10
		Nodular prurigo	18
	Psoriasis		42
	Scabies		2
	Morphoea		7
	Urticaria		2
	Discoid Lupus erythematosus		9
	Hematoma		2
	Cutaneous leishmaniasis		2
	Rash		2
	Rosacea		2
	Postinflammatory hyperpigmentation		13
	Telangiectasia		9

Abbreviations: BCC, basal cell carcinoma; SCC, squamous cell carcinoma.

acquired spectrum was recorded using multiple accumulations of a 300 ms integration time amounting to a total measurement time of 10 seconds. The measurement protocol considered the acquisition of 3 individual spectra with the probe positioned on the suspected lesion and 3 spectra on the adjacent skin (no lesion) as control measurement. The aim of the control measurement was to subtract this from the lesion measurement. However, the control spectra were not used in the data analysis because the classification accuracy of subtracted lesion–control was better when only the lesion spectra were used. The 3 spectra of each lesion measured were averaged for analysis. The spectral range of our recordings ranged between 700 and 1800 cm^{-1}.

The spectra were preprocessed to remove fluorescence and scattering background using a MATLAB asymmetric least squares algorithm (Eilers and Perfect Smoother, 2003); the interference signals arising from the fiber optic probe were removed using an in-house custom-developed algorithm using the MATLAB programming environment. The influence of the detector efficiency on spectral intensities was corrected using a National Insititute of Standards and Technology standard (Choquette et al, 2007). Before classification, the spectra were preprocessed using a Savitsky–Golay smoothing algorithm with 11 adjacent points and a second-order polynomial with area normalization.

Sample collection

Raman spectra were recorded from volunteers at 3 different sites: an advertised, at free university research clinic, and by opportunistic recruitment at both a general dermatology and a plastic surgery clinic. The very experienced dermatologist identified clinically and unambiguously benign lesions and inflammatory dermatoses for

Table 5. Summary of the Different Ethnicity and Lesion Types in Each of the 5 Data Subsets (Calibration + Validation)

Dataset	Total Numbers of Lesions	European	Asian (Indian, Sri Lankan, Filipino, Chinese, Thai, Malaysian)	Māori/Pacific Peoples (Cook Island Māori, Samoan, Tongan)	Latin American, African, and Middle Eastern	Percentage Lesion Type
I	Benign lesions	404	130	41	28	82.9%
	Malignant lesions	118	3	2	1	17.1%
	% Ethnicity	71.8%	18.3%	5.9%	4.0%	
II	Benign lesions	404	130	41	28	69.5%
	Malignant lesions	118	3	2	1	14.3%
	Inflammatory dermatoses	71	17	49	3	16.2%
	% Ethnicity	68.4%	17.3%	10.6%	3.7%	
III	Malignant lesions	118	3	2	1	47.0%
	Inflammatory dermatoses	71	17	49	3	53.0%
	% Ethnicity	71.6%	7.6%	19.3%	1.5%	
IV	BCC	61	1	1	1	51.6%
	Melanoma	24	1	1	1	21.8%
	SCC	33	-	-	-	26.6%
	% Ethnicity	95.2%	1.6%	1.6%	1.6%	
V	Benign lesions	107	2	11	0	47.1%
	Malignant lesions	120	4	2	2	52.9%
	% Ethnicity	91.5%	2.4%	5.2%	0.8%	

Table 6. CV Results and Variance Explained by the PCs Chosen for Training Sets for Data Subsets I–V

Data Subset	Latent Variable Number	X Cumulative Variance Captured (%)	CV Classification Error Average	RMSECV	Data Subset	Latent Variable Number	X Cumulative Variance Captured (%)	CV Classification Error Average	RMSECV
I (i)	1	53.13	0.3537	0.3511	I (ii)	1	32.19	0.5161	0.2529
	2	64.22	0.2859	0.3393		2	63.45	0.4275	0.2512
	3	71.69	0.2353	0.3283		3	73.61	0.3597	0.2476
	4	78.89	0.2231	0.3208		4	81.84	0.3499	0.2494
	5	84.74	0.2067	0.3162		5	86.24	0.3554	0.2474
	6	86.67	0.1997	0.31		6	89.35	0.3084	0.2452
	7	88.65	0.1979	0.3082		7	92.08	0.2836	0.2419
	8	91.89	0.196	0.3052		8	92.65	0.2667	0.2126
	9	93.23	0.1705	0.3024	III	1	52.06	0.3798	0.4925
	10	93.87	0.171	0.2996		2	60.74	0.3375	0.4696
	11	94.89	0.1682	0.3		3	67.94	0.2881	0.4448
II	1	53.47	0.3703	0.3332		4	77.87	0.2924	0.4442
	2	64.07	0.3121	0.324		5	82.15	0.2301	0.4274
	3	71.31	0.2388	0.3154		6	85.99	0.2114	0.4165
	4	78.54	0.2383	0.3106		7	88.55	0.2044	0.4022
	5	83.52	0.2136	0.3074		8	91.16	0.2011	0.397
	6	86.5	0.2146	0.304		9	91.92	0.1981	0.3827
	7	90.49	0.2182	0.3023		10	93.31	0.1832	0.3681
	8	92.13	0.2189	0.2982		11	94.61	0.1674	0.3616
	9	93.11	0.1971	0.2942		12	95.57	0.1525	0.3596
	10	93.58	0.1973	0.2889		13	95.91	0.1489	0.3558
	11	94.63	0.2008	0.2884	V	1	52.06	0.3872	0.4924
IV	1	44.14	0.2964	0.3538		2	60.74	0.3227	0.4693
	2	64.4	0.164	0.324		3	67.94	0.2977	0.4433
	3	73.3	0.176	0.3095		4	77.87	0.2851	0.4465
	4	78.99	0.201	0.3273		5	82.15	0.2354	0.4342
	5	82.57	0.164	0.331		6	85.99	0.2274	0.4244
	6	85.67	0.1455	0.3359		7	88.55	0.2179	0.4123
	7	87.09	0.126	0.3466		8	91.16	0.2233	0.4084
	8	90.19	0.139	0.3318		9	91.92	0.2064	0.3933
	9	92.43	0.1205	0.3178		10	93.31	0.1894	0.3802
	10	93.53	0.139	0.3062		11	94.61	0.1671	0.3714
	11	94.84	0.139	0.3013		12	95.57	0.1638	0.3694
						13	95.91	0.1597	0.3691

Abbreviations: CV, cross-validation; PC, principal component; RMSECV, root mean squared error of cross-validation.

Raman spectroscopy. The majority of these diagnoses did not have routine histological confirmation, reflecting usual clinical practice. Those that were identified by the dermatologist as being malignant were referred for biopsy for histology testing. At the plastic surgery clinic, 2 plastic surgeons again triaged the patients who had been referred by their general practitioners as having likely or skin cancers (melanoma, BCC, and SCC). After in situ Raman spectroscopy measurements of these lesions, those selected for biopsy were sent for histology confirmation as malignant or benign after the excision.

In total, 1,296 benign and cancer lesions were measured from 320 patients. A total of 78% of the volunteers were European. Of the European samples, 41% were recognized as benign nevi. Of note, up to 20 different nevi were measured from different parts of the body on a number of the participants. The numbers of benign nevi in the dataset were therefore reduced by randomly selecting a maximum of 4 benign lesions on different areas of the body from any individual participant. The total dataset was thus reduced to 867

spectra that comprised 70% European (closer to the 70.2% European demographic in New Zealand [Statistics New Zealand, 2018]). The ethnicity breakdown of the total dataset of 867 spectra was 70% European; 16% Asian; 6% Māori; 5% Pacific people; and 4% Middle Eastern, Latin American, and African patients.

The 124 malignant lesion samples included 27 melanomas (lentiginous melanoma, spitzoid melanoma, melanoma in situ, superficial spreading melanoma, metastatic melanoma, and acral melanoma), 64 BCCs (in situ, nodular, superficial spreading, multifocal, and infiltrative), and 33 SCCs (in situ, acantholytic, infiltrative, and poorly/moderately/well-differentiated). The 743 benign skin conditions included a variety of both melanocytic and nonmelanocytic lesions (603 in total) and 140 inflammatory dermatoses. Melanocytic lesions included nevi (congenital, compound, blue, intradermal, lentiginous, junctional, halo, pedunculated, and speckled lentiginous nevi), lentigo, café au lait, ephelides, and solar lentigo. Nonmelanocytic benign lesions included actinic keratosis

and SK, linear epidermal nevus, keratoacanthomas, sebaceous hyperplasia, haemangioma, Becker nevus, fibroxanthoma, fibrous papule, dermatofibroma, scar tissue, vascular malformation, scar tissue, plantar fibromatosis, myxoid and milia cysts, keratosis pilaris, cherry angioma, acrochordon, and accessory nipple. The inflammatory dermatoses included different eczema types (atopic eczema, discoid eczema, dermatitis not otherwise specified, nodular prurigo), psoriasis, morphea, urticaria, and discoid lupus erythematosus. The numbers of benign lesions are more than 3 times greater than the numbers of malignant lesions; however, the intent was to include as wide a variety of benign lesions as possible in the dataset that could be encountered in a general clinic. Because it was not possible to biopsy every benign lesion for histology testing, an additional dataset subset V that comprised only those lesions with histology test results (benign and malignant) was also included in the study.

The data subsets I—V mentioned earlier were each split into a training set and a separate validation set in the ratio ~85%: ~15%. Care was taken to ensure that for each data subset, the training and validation sets did not share any data from the same patient. The validation sets therefore represented a separate set of spectra and from different patients.

The different types of lesions and inflammatory dermatoses are listed in Table 4. The ethnicities and lesion types in the 5 different datasets are summarized in Table 5. Classification models were created for 5 different data subsets to aid clinicians in the diagnosis of different lesion types in a routine clinic.

Data analysis

The classification of different lesion types was performed using PLS-DA, chosen for its suitability for classifying Raman spectra of biological tissue. The very large number of variables in Raman spectra can hinder the classification, especially when the number of samples is limited, so the dimensionality reduction of the PLS method is ideal. Cross-validation was performed on the training sets using the random subset cross-validation procedure. The data training set was split randomly into 10 smaller groups; one group was then subjected to prediction by a calibration model drawn up using the other 9 groups and, this was repeated for each group. Five iterations of this were procedure performed on 5 different groups of 10 data groups to give an average root mean square error of the cross-validation for the 5 different iterations. The number of PCs or latent variables was selected as that number yielding the minimum root mean square error of the cross-validation. The cross-validation results and variance explained by the PCs chosen for each training set in data subsets I—V are listed in Table 6.

The diagnostic accuracy of each classification model was evaluated in terms of the AUC of the ROC and calculated using MedCalc (version 22.001, MedCalc Software). Although other criteria such as sensitivity, specificity, positive predictive values, and negative predictive values are often used, they can result in noticeable trade offs between themselves (Trevethan, 2017). The ROC plots the cross-validated sensitivity (1 − specificity) for all possible threshold values that could be defined for predicting a lesion as malignant or benign and thus enables visualization of this trade off between sensitivity (plotted on the y-axis) and 1 − specificity (plotted on the x-axis). The threshold for defining a positive test result (cancer vs benign) varies from most stringent (100% sensitivity) to least stringent (100% specificity) (Kraemer, 1992; Hulley and Cummings, 1988). Generally, the optimum threshold proposed by PLS-DA corresponds with the probability of being in the malignant or benign class equal to 50%. However, it is desirable for clinicians to select the threshold that will ensure that all malignant samples are correctly classified as malignant (ie, 100% sensitivity). In this study, the specificities obtained at the sensitivity levels 100, 95, and 90% were determined for each dataset.

ETHICS STATEMENT

Ethical approval was obtained from the University of Auckland Human Participants Ethics Committee (reference number UAHPEC20101), the Health and Disability Ethics Committee (reference number 8129), and the Counties Manukau Health research office (reference number 850). All participants signed informed consent.

DATA AVAILABILITY STATEMENT

Data underlying the results presented in this paper may be obtained from the corresponding author upon reasonable request.

ORCIDs
Michel Nieuwoudt: http://orcid.org/0000-0003-4963-412X
Paul Jarrett: http://orcid.org/0000-0001-5030-8896
Ira Mautner: http://orcid.org/0000-0002-5801-1103
Hannah Matthews: http://orcid.org/0000-0003-4542-1194
Michelle Locke: http://orcid.org/0000-0002-8388-6390
M. Cather Simpson: http://orcid.org/0000-0001-9624-4947
Claude Aguergaray: http://orcid.org/0000-0001-5300-9920
Hannah Holtkamp: http://orcid.org/0000-003-2675-2537
Thom Minnee: http://orcid.org/0000-0002-8773-2110
Marco Bonesi: http://orcid.org/0000-0001-5194-8713
Brydon Burnett: http://orcid.org/0000-0002-6944-8159

AUTHOR CONTRIBUTIONS
Conceptualization: MN, MCS, PJ, ML; Data Curation: MN, HM, HH, BB, TM; Formal Analysis: MN, BB, HM, PJ; Funding Acquisition: MN, MCS; Investigation: MN, PJ, HM, HH, ML; Methodology: MN, HM, HH, PJ, ML, MB, CA; Project Administration: MN, MCS; Resources: MCS, PJ, ML; Software: BB, MB, CA, TM; Supervision: MN, MCS; Validation: MN, HM; Visualization: MN; Writing − Original Draft Preparation: MN

CONFLICT OF INTEREST
The authors state no conflict of interest.

REFERENCES

Baclig AC, Bakker Schut TC, O'Regan GM, Irvine AD, McLean WHI, Puppels GJ, et al. Possibilities for human skin characterization based on strongly reduced Raman spectroscopic information. J Raman Spectrosc 2013;44:340−5.

Barroso EM, Smits RW, Bakker Schut TC, ten Hove I, Hardillo JA, Wolvius EB, et al. Discrimination between oral cancer and healthy tissue based on water content determined by Raman spectroscopy. Anal Chem 2015;87:2419−26.

Bratchenko IA, Artemyev DN, Myakinin OO, Khristoforova YA, Moryatov AA, Kozlov SV, et al. Combined Raman and autofluorescence ex vivo diagnostics of skin cancer in near-infrared and visible regions. J Biomed Opt 2017;22:027005.

Bratchenko IA, Bratchenko LA, Moryatov AA, Khristoforova YA, Artemyev DN, Myakinin OO, et al. In vivo diagnosis of skin cancer with a portable Raman spectroscopic device. Exp Dermatol 2021;30:652−63.

Brauchle E, Noor S, Holtorf E, Garbe C, Schenke-Layland K, Busch C. Raman spectroscopy as an analytical tool for melanoma research. Clin Exp Dermatol 2014;39:636−45.

Brozek-Pluska B, Musial J, Kordek R, Bailo E, Dieing T, Abramczyk H. Raman spectroscopy and imaging: applications in human breast cancer diagnosis. Analyst 2012;137:3773−80.

Cals FL, Bakker Schut TC, Hardillo JA, Baatenburg de Jong RJ, Koljenović S, Puppels GJ. Investigation of the potential of Raman spectroscopy for oral cancer detection in surgical margins. Lab Invest 2015;95:1186−96.

Carvalho LF, Bonnier F, O'Callaghan K, O'Sullivan J, Flint S, Byrne HJ, et al. Raman micro-spectroscopy for rapid screening of oral squamous cell carcinoma. Exp Mol Pathol 2015;98:502−9.

Choi J, Choo J, Chung H, Gweon DG, Park J, Kim HJ, et al. Direct observation of spectral differences between normal and basal cell carcinoma (BCC) tissues using confocal Raman microscopy. Biopolymers 2005;77:264−72.

Choquette SJ, Etz ES, Hurst WS, Blackburn DH, Leigh SD. Relative intensity correction standard for Raman spectroscopy: NIST SRMs 2241 through 2243 for 785 nm, 532 nm, and 488 nm/514.5 nm excitation. Appl Spectrosc 2007;61:117−29.

Eilers PHC. A perfect smoother. Anal Chem 2003;75:3299−304.

Essendoubi M, Gobinet C, Reynaud R, Angiboust JF, Manfait M, Piot O. Human skin penetration of hyaluronic acid of different molecular weights as probed by Raman spectroscopy. Skin Res Technol 2016;22:55−62.

Feng XF, Fox MC, Reichenberg JS, Lopes FCPS, Sebastian KR, Markey MK, et al. Biophysical basis of skin cancer margin assessment using Raman spectroscopy. Biomed Opt Express 2018;10:104−18.

Fox SA, Shanblatt AA, Beckman H, Strasswimmer J, Terentis AC. Raman spectroscopy differentiates squamous cell carcinoma (SCC) from normal skin following treatment with a high-powered CO2 laser. Lasers Surg Med 2014;46:757−72.

Gniadecka M, Nielsen OF, Wulf HC. Water content and structure in malignant and benign skin tumours. J Mol Struct 2003;661−662:405−10.

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;122:443−9.

Gniadecka M, Nielsen OF, Wessel S, Heidenheim M, Christensen DH, Wulf HC. Water and protein structure in photoaged and chronically aged skin. J Invest Dermatol 1998;111:1129−33.

Hammody Z, Sahu RK, Mordechai S, Cagnano E, Argov S. Characterization of malignant melanoma using vibrational spectroscopy. ScientificWorldJournal 2005;5:173−82.

Heal CF, Raasch BA, Buettner PG, Weedon D. Accuracy of clinical diagnosis of skin lesions. Br J Dermatol 2008;159:661−8.

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;107:1047−52.

Huang Z, Zeng H, Hamzavi I, McLean DI, Lui H. Rapid near-infrared Raman spectroscopy system for real-time in vivo skin measurements. Opt Lett 2001;26:1782−4.

Hulley S, Cummings S. Designing clinical research. Baltimore, MD: Williams & Wilkins; 1988.

Khristoforova YA, Bratchenko IA, Myakinin OO, Artemyev DN, Moryatov AA, Orlov AE, et al. Portable spectroscopic system for in vivo skin neoplasms diagnostics by Raman and autofluorescence analysis. J Biophotonics 2019;12:e201800400.

Kong K, Rowlands CJ, Varma S, Perkins W, Leach IH, Koloydenko AA, et al. Diagnosis of tumors during tissue-conserving surgery with integrated autofluorescence and Raman scattering microscopy. Proc Natl Acad Sci USA 2013;110:15189−94.

Kourkoumelis N, Balatsoukas I, Moulia V, Elka A, Gaitanis G, Bassukas ID. Advances in the in vivo Raman spectroscopy of malignant skin tumors using portable instrumentation. Int J Mol Sci 2015;16:14554−70.

Kraemer H. Evaluating medical tests: objective and quantitative guidelines. Thousand Oaks, CA: USA Sage Publications; 1992.

Larraona-Puy M, Ghita A, Zoladek A, Perkins W, Varma S, Leach IH, et al. Development of Raman microspectroscopy for automated detection and imaging of basal cell carcinoma. J Biomed Opt 2009;14:054031.

Latka I, Dochow S, Krafft C, Dietzek B, Popp J. Fiber optic probes for linear and nonlinear Raman applications - current trends and future development. Laser Photonics Rev 2013;7:698−731.

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

Lieber CA, Majumder SK, Ellis DL, Billheimer DD, Mahadevan-Jansen A. In vivo nonmelanoma skin cancer diagnosis using Raman microspectroscopy. Lasers Surg Med 2008b;40:461−7.

Lim L, Nichols B, Migden MR, Rajaram N, Reichenberg JS, Markey MK, et al. Clinical study of noninvasive in vivo melanoma and nonmelanoma skin cancers using multimodal spectral diagnosis. J Biomed Opt 2014;19:117003.

Lui H, Zhao J, McLean D, Zeng H. Real-time Raman spectroscopy for in vivo skin cancer diagnosis. Cancer Res 2012;72:2491−500.

MacKenzie-Wood AR, Milton GW, de Launey JW. Melanoma: accuracy of clinical diagnosis. Australas J Dermatol 1998;39:31−3.

McArdle L, Rafferty M, Maelandsmo GM, Bergin O, Farr CJ, Dervan PA, et al. Protein tyrosine phosphatase genes downregulated in melanoma. J Invest Dermatol 2001;117:1255−60.

Morton CA, Mackie RM. Clinical accuracy of the diagnosis of cutaneous malignant melanoma. Br J Dermatol 1998;138:283−7.

Motz JT, Hunter M, Galindo LH, Gardecki JA, Kramer JR, Dasari RR, et al. Optical fiber probe for biomedical Raman spectroscopy. Appl Opt 2004;43:542−54.

Movasaghi Z, Rehman S, Rehman IU. Raman spectroscopy of biological tissues. Appl Spectrosc Rev 2007;42:493−541.

Nguyen TT, Gobinet C, Feru J, Pasco SB, Manfait M, Piot O. Characterization of type I and IV collagens by Raman microspectroscopy: identification of spectral markers of the dermo-epidermal junction. Spectrosc Int J 2012;27:421−7.

Nijssen A, Maquelin K, Santos LF, Caspers PJ, Bakker Schut TC, den Hollander JC, et al. Discriminating basal cell carcinoma from perilesional skin using high wave-number Raman spectroscopy. J Biomed Opt 2007;12:034004.

Santos IP, van Doorn R, Caspers PJ, Bakker Schut TC, Barroso EM, Nijsten TEC, et al. Improving clinical diagnosis of early-stage cutaneous melanoma based on Raman spectroscopy. Br J Cancer 2018;119:1339−46.

Puppels GJ, Garritsen HSP, Segers-Nolten GMJ, de Mul FFM, Greve J. Raman microspectroscopic approach to the study of human granulocytes. Biophys J 1991;60:1046−56.

Santos IP, Barroso EM, Bakker Schut TC, Caspers PJ, van Lanschot CGF, Choi DH, et al. Raman spectroscopy for cancer detection and cancer surgery guidance: translation to the clinics. Analyst 2017;142:3025−47.

P Santos IP, Caspers PJ, Bakker Schut TC, van Doorn R, Noordhoek Hegt V, Koljenović S, et al. Raman spectroscopic characterization of melanoma and benign melanocytic lesions suspected of melanoma using high-wavenumber Raman spectroscopy. Anal Chem 2016;88:7683−8.

Schallreuter KU, Wood JM. The importance of L-phenylalanine transport and its autocrine turnover to L-tyrosine for melanogenesis in human epidermal melanocytes. Biochem Biophys Res Commun 1999;262:423−8.

Sharma M, Marple E, Reichenberg J, Tunnell JW. Design and characterization of a novel multimodal fiber-optic probe and spectroscopy system for skin cancer applications. Rev Sci Instrum 2014;85:083101.

Silveira FL, Pacheco MT, Bodanese B, Pasqualucci CA, Zângaro RA, Silveira L Jr. Discrimination of non-melanoma skin lesions from non-tumor human skin tissues in vivo using Raman spectroscopy and multivariate statistics. Lasers Surg Med 2015;47:6−16.

Statistics New Zealand. Ethnic Population Estimates at 30 Jine. 2018. https://www.stats.govt.nz/information-releases/estimated-resident-population-2018-base-at-30-june-2018#ethnic. (accessed October 26, 2022).

Stone N, Kendall C, Smith J, Crow P, Barr H. Raman spectroscopy for identification of epithelial cancers. Faraday Discuss 2004;126:141−57. discussion 169−83.

Trevethan R. Sensitivity, specificity, and predictive values: foundations, pliabilities, and pitfalls in research and practice. Front Public Health 2017;5: 307.

Ying GS, Maguire MG, Glynn RJ, Rosner B. Calculating sensitivity, specificity, and predictive values for correlated eye data. Invest Ophthalmol Vis Sci 2020;61:29.

Zhang Q, Andrew Chan KL, Zhang G, Gillece T, Senak L, Moore DJ, et al. Raman microspectroscopic and dynamic vapor sorption characterization of hydration in collagen and dermal tissue. Biopolymers 2011;95:607−15.