


## LETTER

# Supervised learning methods for the recognition of melanoma cell lines through the analysis of their Raman spectra

Enrico Baria<sup>1,2\*</sup>  | Riccardo Cicchi<sup>2,3</sup> | Francesca Malentacchi<sup>4</sup> | Irene Mancini<sup>4</sup> | Pamela Pinzani<sup>4</sup> | Marco Pazzagli<sup>4</sup> | Francesco S. Pavone<sup>1,2,3</sup>

<sup>1</sup>Department of Physics, University of Florence, Sesto Fiorentino, Italy

<sup>2</sup>European Laboratory for Non-Linear Spectroscopy, University of Florence, Sesto Fiorentino, Italy

<sup>3</sup>National Institute of Optics, National Research Council, Florence, Italy

<sup>4</sup>Department of biomedical, experimental, and clinical sciences “Mario Serio”, University of Florence, Florence, Italy

## \*Correspondence

Enrico Baria, Department of Physics, University of Florence, 50019 Sesto Fiorentino, Italy.  
Email: baria@lens.unifi.it

## Funding information

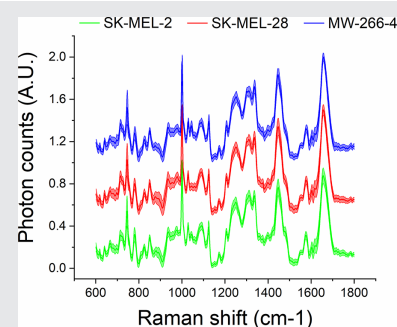
Ente Cassa di Risparmio di Firenze; Horizon 2020 Framework Programme, Grant/Award Number: 654148 Laserlab-Europe; Ministero dell'Istruzione, dell'Università e della Ricerca, Grant/Award Number: NANOMAX

## Abstract

Malignant melanoma is an aggressive form of skin cancer, which develops from the genetic mutations of melanocytes – the most frequent involving *BRAF* and *NRAS* genes. The choice and the effectiveness of the therapeutic approach depend on tumour mutation; therefore, its assessment is of paramount importance. Current methods for mutation analysis are destructive and take a long time; instead, Raman spectroscopy could provide a fast, label-free and non-destructive alternative. In this study, confocal Raman microscopy has been used for examining three in vitro melanoma cell lines, harbouring different molecular profiles and, in particular, specific *BRAF* and *NRAS* driver mutations. The molecular information obtained from Raman spectra has served for developing two alternative classification algorithms based on linear discriminant analysis and artificial neural network. Both methods provide high accuracy ( $\geq 90\%$ ) in discriminating all cell types, suggesting that Raman spectroscopy may be an effective tool for detecting molecular differences between melanoma mutations.

## KEYWORDS

cells, melanoma, neural network, Raman spectroscopy, supervised learning



## 1 | INTRODUCTION

Malignant melanoma (MM) is the most dangerous form of skin cancer, due to its strong aggressiveness and the high mortality rate in case of late diagnosis: estimates for 5-year survival rates for patients with a metastatic disease

range from 25% to 5% [1, 2]. MM develops when unrepaired DNA damages to skin cells – most often caused by ultraviolet radiation from sunshine – trigger mutations (i.e., genetic defects) that lead the melanocytes multiplying rapidly and form a malignant tumour [3]. Clinical treatment of MM consists of a combination of surgery, traditional cytotoxic chemotherapy, targeted therapies, and immune-based therapies. The response to both conventional chemotherapy and targeted agents depends on the specific mutation involved in the

**Abbreviations:** ANN, artificial neural network; LDA, linear discriminant analysis; MM, malignant melanoma; PCA, principal component analysis.

carcinogenesis. In particular, the most frequent oncogenic mutations involve *BRAF* and *NRAS* genes (approximately 50% and 20% of all melanomas, respectively [4, 5]); mutations in more than one of these genes are seldom found concurrently in the same tumour. In fact, *BRAF* and *RAS* encoded proteins are components of the MAPK (mitogen-activated protein kinases) pathway that promotes the activation of transcription factors important for cell growth, proliferation, and survival. Several mutations in these genes can result in MAPK pathway hyperactivation, thus driving the development of melanoma.

Assessing the molecular profiles affected by driver mutations is extremely important for proper clinical treatment of the disease. Current methods of mutation analysis are based on DNA sequencing [6] and other polymerase chain reaction (PCR [7]) based assays; however, these techniques are destructive and time-consuming. The development of accurate approaches for the recognition of cancer cell lines and for testing the purity of in vitro cultures are strongly desired as well, due to their use as models to study tumour phenomena.

In this context, Raman spectroscopy is a possible alternative to the standard methods. It provides a fast, non-destructive, label-free optical tool for studying living cells [8, 9] and, specifically, for detecting the cancer ones [10–12]. In fact, it could be used for investigating the molecular composition of different cell types, highlighting their differences.

In this study, we used confocal Raman microscopy for examining in vitro melanoma cells with specific mutations of either *BRAF* or *NRAS* genes. The presented results validated a previous work (Brauchle et al. [10], 2014), provided a biochemical interpretation of the observed differences between MM mutations, and expanded the spectral analysis: the information provided by the recorded Raman dataset was used for developing two alternative classification algorithms based, respectively, on linear discriminant analysis (LDA) and artificial neural network (ANN). Finally, we evaluated and compared the diagnostic accuracy of these two methods when applied to the analysis of MM cell spectra, obtaining ~93% and ~97% overall success rates for LDA and ANN, respectively.

## 2 | MATERIALS AND METHODS

### 2.1 | Sample preparation

Three in vitro cell cultures of human melanoma cell lines were analysed, each one belonging to a different genotype: SK-MEL-28 (homozygous mutant for classical variant p.Val600Glu, or V600E, in the *BRAF* gene), SK-MEL-2 (homozygous mutant for *NRAS* p.Gln61Arg, or Q61R) and

MW-266-4 (without classical mutations in both *BRAF* and *NRAS* genes, but characterised by the p.Val600Asp, or V600D, variant [13] – a rare *BRAF* mutation in heterozygous status). SK-MEL-28 and MW-266-4 cell lines were expanded from clones of the American Type Culture Collection (ATCC), while SK-MEL-2 was obtained from CLS cell lines service GmbH (Eppelheim, Germany).

All the cell lines were grown in incubator at 37°C with 5% of CO<sub>2</sub> using DMEM +10% fetal bovine serum (FBS) (Euroclone, Milano, Italy), on glass-bottom dishes (MatTek, Ashland, Massachusetts). The culture medium was removed and substituted with phosphate-buffered saline (PBS) solution immediately before Raman measurements.

### 2.2 | Experimental setup

A commercial setup was used for Raman measurements: XploRA INV (Horiba, Kyoto, Japan), an inverted, confocal Raman microscope. The microscope consists of a continuous-wave laser emitting 20 to 25 mW at 532 nm, a dual mechanical/optical scanning system for delivering the excitation light on various portions of the sample, a Plan-Apochromat 60× objective (Nikon, Tokyo, Japan; NA = 1.27, WD = 0.17 mm), a spectrograph equipped with a CCD camera and four gratings with different spacing (600, 1200, 1800, and 2400 lines/mm). The microscope is remotely controlled via PC software (LabSpec6, Horiba, Kyoto, Japan). In addition, two white-light sources allow for wide field epi- and trans-illumination imaging, respectively.

Raman measurements were performed in epi-detection configuration. The inverted configuration of the system is particularly suited for studying cell cultures.

### 2.3 | Raman spectra acquisition

White-light illumination was used to image cell cultures and to identify cell nuclei. In order to acquire a Raman spectrum from each nucleus under investigation, a 5 µm size radial area centred on the nucleus was scanned for 120 seconds with the excitation laser (which has 1.5 and 6.5 µm radial and axial spots, respectively). The scattered light was backward collected and detected between 300 and 1900 cm<sup>-1</sup> after being diffracted by the 1800 lines/mm grating. The 50 Raman spectra were recorded for each cell type (150 in total).

### 2.4 | Data preprocessing

We observed the presence of unwanted spectral contributions in the recorded data as, for example, the broad

Raman bands generated by the glass [14] bottom of MatTek dishes. Therefore, the acquired spectra were processed for removing both glass and PBS signals, and the fluorescence emission from the sample. Specifically:

- the normalised Raman spectrum recorded from the PBS solution inside the MatTek was subtracted from all the normalised spectra recorded from melanoma cell nuclei;
- the spectral range of each cell spectrum was restricted between  $600\text{ cm}^{-1}$  and  $1800\text{ cm}^{-1}$  (i.e., the so-called “fingerprint region”);
- an automated iterative routine (Vancouver Raman Algorithm [15]) was used for fitting the residual fluorescence spectral contribution to a 5th order polynomial function, and for subtracting it from each cell spectrum.

The resulting Raman spectra were normalised and analysed through Principal Component Analysis (PCA), a well-known method [16] for extracting the most relevant information from multidimensional datasets. We initially selected the first 15 Principal Components (PCs), corresponding to 89.7% of the total variance in our spectral dataset. For each PC, we derived the Receiver Operating Characteristic (ROC) curve of each cell type and measured the corresponding area under the curve (auROCc), which can be used to estimate the PC discriminatory power. Finally, the 15 PCs were ranked according to their average auROCc value for all cell types.

## 2.5 | Data analysis

### 2.5.1 | Analytical methods

The loadings obtained from PCA were separately analysed through two different methods: LDA and ANN. Both techniques can be combined with PCA [17, 18] in order to classify spectral datasets. The former, LDA, uses a linear classifier to separate PCA loadings of multiple classes (here based on one-vs-rest approach) and maximise the average separation between them. In this case, a score is assigned to each cell nucleus, and a threshold is automatically chosen for separating scores belonging to different cell lines. The latter, ANN, uses the loadings as inputs for a network of simple, highly interconnected processing elements (called *nodes* or *neurons*); such network is trained to classify the input data through an iterative routine based on supervised learning. In particular, the neural network was trained to output a different numerical code – or “target” – for each cell line (1-0-0 for SK-MEL-2, 0-1-0 for SK-MEL-28, and 0-0-1 for MW-266-4).

LDA and ANN were implemented through two built-in routines in MATLAB (MathWorks, Natick, MA, US): Classification Learner (CL) and Neural Network Pattern Recognition (NNPR). A basic description of these tools is provided in the Supporting Information.

### 2.5.2 | Cross-validation

5-fold cross-validation was applied to both analyses. The dataset was partitioned in 5 subgroups, each one consisting of 30 spectra (10 from each cell line). During each round of the cross-validation procedure, 4 subgroups (“learning set”, totalling 120 spectra) were pooled together for training the LDA/ANN classification algorithm through supervised learning, while the last one (“testing set”) was used to evaluate its accuracy. All subgroups were rotated between these two sets during the 5 rounds of learning and testing. In order to better compare the performances of LDA and ANN, all subgroups and rotations were identical for both analyses.

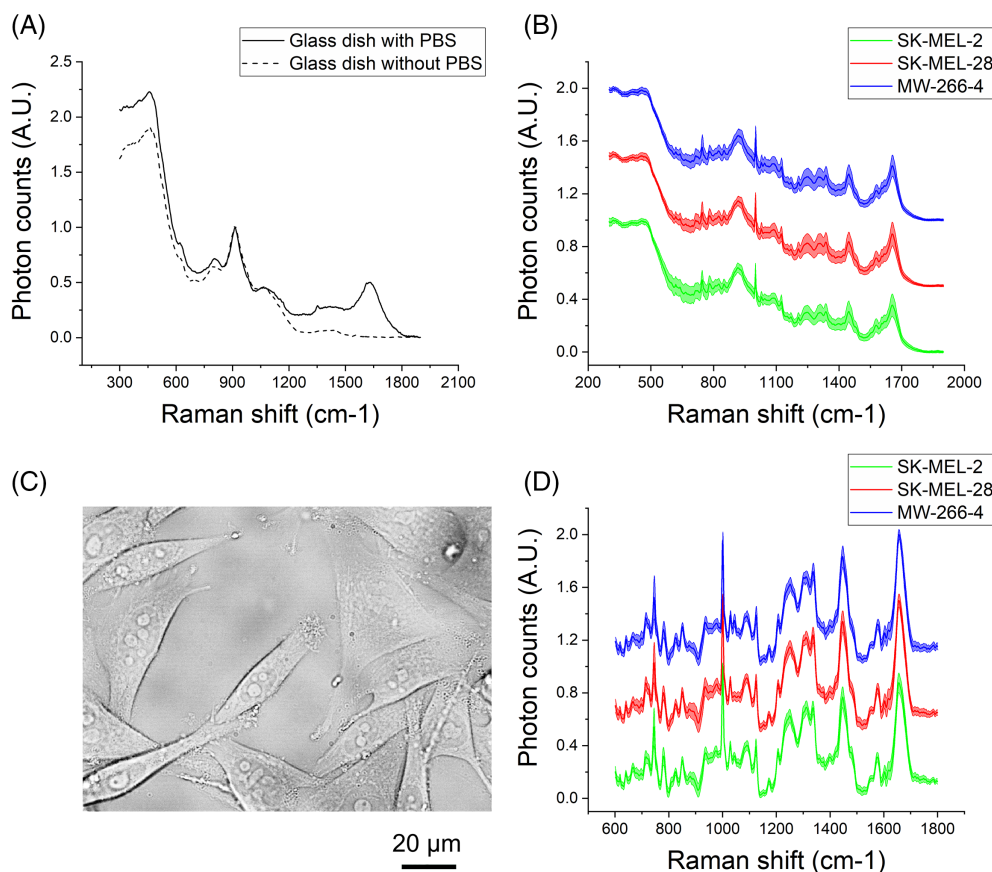
### 2.5.3 | PCs selection

In order to reduce the risk of overfitting the data, the number of PCs to be included in the final analysis was chosen based on normalised root-mean-square error (NRMSE) estimation. The loadings of the most discriminant PC were analysed through both LDA and ANN, hence the NRMSE averaged across all five rounds of cross-validation (NRMSE-CV) was calculated; the same procedure was performed by including an incremental number of PCs, sorted according to their discriminatory power. Finally, the NRMSE-CV minimum value indicated which PCs had to be included, while the remaining ones were discarded. Additional information about this analysis are provided in the Supporting Information.

## 3 | RESULTS AND DISCUSSION

### 3.1 | Spectral preprocessing

As explained in Section 2.4 glass is known [14] to give a spectral contribution to Raman spectra for 532 nm excitation wavelength. Thus, we recorded the spectrum of the glass-bottom dish with and without PBS. Both spectra were normalised to the intensity of the glass band at  $915\text{ cm}^{-1}$  in order to better recognise the differences due to PBS presence/absence (Figure 1(A)): glass has major Raman contributions in the  $300\text{--}600\text{ cm}^{-1}$  and  $800\text{--}1200\text{ cm}^{-1}$  spectral ranges, while PBS generates a



**FIGURE 1** A, comparison between the Raman spectra acquired on a glass-bottom dish with and without PBS solution. B, mean “raw” Raman spectra  $\pm$  standard deviations for SK-MEL-2 (green), SK-MEL-28 (red) and MW-266-4 (blue) cells. For better comparison, the spectra are vertically shifted. C, wide field epi-illumination images of SK-MEL-2 mutated cells, obtained with a white-light source. D, mean preprocessed Raman spectra  $\pm$  standard deviations of SK-MEL-2, SK-MEL-28 and MW-266-4 cells obtained after removing glass, PBS background, and fluorescence baseline from the raw acquired data. Again, the spectra are vertically shifted. PBS, phosphate-buffered saline

large peak around  $1630\text{ cm}^{-1}$ . Due to such contributions, “raw” cell spectra (Figure 1(B)) are characterised by a distinct background signal originating from both these substances, which decreases the signal-to-noise ratio (SNR) of cell Raman peaks. Therefore, as described in Section 2.4, the glass/PBS background spectrum was subtracted from cell spectra after normalising each spectrum to its maximum intensity. Following such operation, the resulting spectra of melanoma cells were further processed for removing the remaining fluorescence baseline. Figure 1(D) shows the final output of such process, that is, the preprocessed Raman spectra of the three cell lines. The presented method for preprocessing raw data allowed reducing variability between cells of the same type and obtaining higher SNR, which is of paramount importance for detecting the small changes in molecular composition between different cell lines.

### 3.2 | Raman bands assignment

The Raman spectra shown in Figure 1(D) are characterised by typical Raman bands associated with the presence of proteins, DNA and lipids. In fact, the main peaks can be assigned to amino acids (tryptophan:  $750, 1340\text{ cm}^{-1}$ ; tyrosine:  $850\text{ cm}^{-1}$ ; phenylalanine:  $1000,$

$1040\text{ cm}^{-1}$ ; amide I:  $1660\text{ cm}^{-1}$ ; amide III:  $1250, 1300, 1340\text{ cm}^{-1}$ ), nucleic acids ( $780, 1030, 1090, 1170, 1300, 1580\text{ cm}^{-1}$ ) and lipids ( $1300, 1450, 1660\text{ cm}^{-1}$ ). The reported spectra have few observable differences among them (Figure S2), most notably: SK-MEL-28 cells show significantly smaller contribution at  $1040\text{ cm}^{-1}$  (phenylalanine) with respect to the other cell types, while SK-MEL-2 cells have lower peak intensities at  $1450$  and  $1660\text{ cm}^{-1}$  (lipids, amide I) and MW-266-4 at  $780$  and  $1480\text{ cm}^{-1}$  (nucleic acids). These findings suggest that Raman spectra could highlight differences both at the genetic level and in the protein synthesis.

### 3.3 | Cell type classification through PCA-LDA

The observable differences between the three cell types are quite small, indicating that most of the recorded spectral data might be “redundant” for discriminating them. Hence, PCA can be used for extracting only the information needed for recognising tumour cell types. Loadings are usually plotted into 2D graphs, like the one shown in Figure 2(A) and (B): horizontal and vertical axes report values of different loadings, and each data point corresponds to one recorded spectrum. Ideally, by plotting the



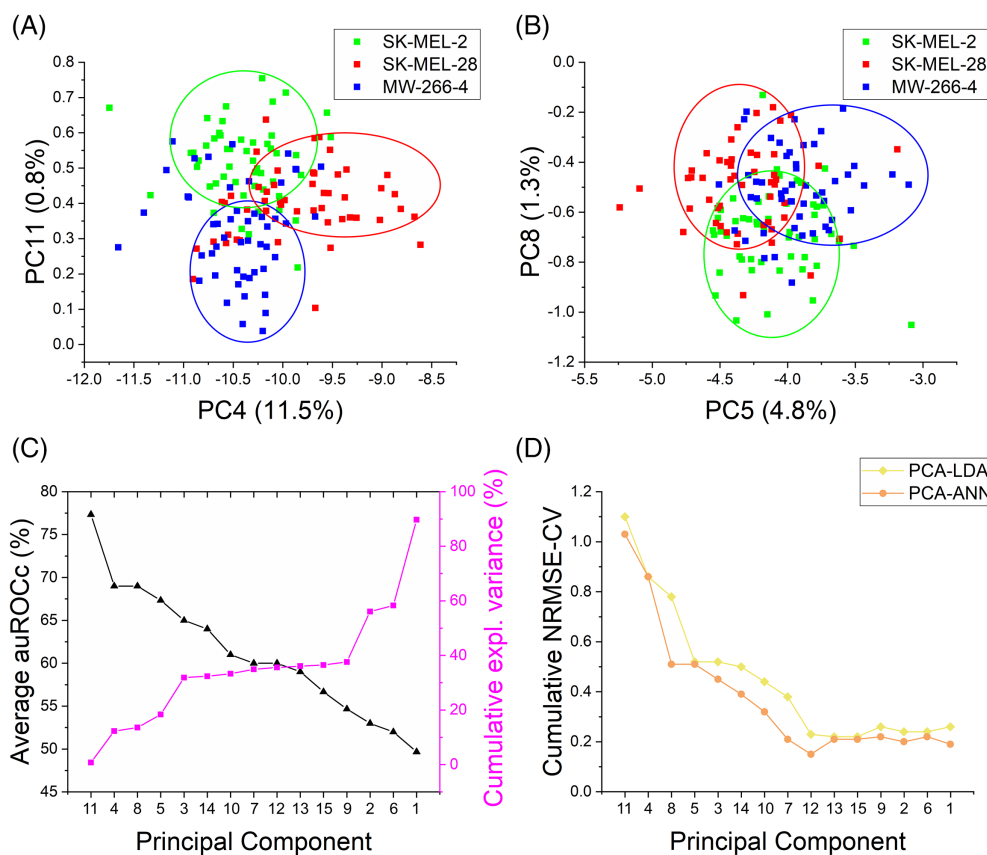
proper loadings, data points belonging to different classes (e.g. different cell types) should flock into spatially separated “clusters”. In this case, however, SK-MEL-2, SK-MEL-28 and MW-266-4 data points cannot be easily separated: discriminating these cell types requires combining the information provided by more than two PCs. As explained in Section 2.5.1, LDA—through the CL app—can be used to generate a scoring system based on the linear separation of multiple PC loadings belonging to different classes. However, using all the available PCs could potentially lead to overfit the “learning set” and poorly classify the “testing set” of each cross-validation round. Therefore, in order to select a proper number of loadings, the NRMSE-CV was estimated for increasing numbers of PCs included in the analysis, starting from the most discriminant PC (PC11) and ending by including the least one (PC1). Figures 2(C) reports the first 15 PCs, ranked from left to right according to their discriminatory power, and Figure 2(D) shows NRMSE-CV trends estimated for both LDA and ANN techniques. The corresponding minima were found by combining the most important 10 and 9 PCs, respectively, which together accounted for ~36% of the total variance. Since the improvement gained by including the 10th PC (PC13) was marginal, we decided to keep only the first nine

loadings also for LDA; the remaining PCs were discarded in order to reduce noise, redundant information and problem dimensionality.

As explained in Section 2.5.2, a 5-fold cross-validation routine was implemented during LDA. The results of all five rounds are reported in Table 1: 92.7% of all testing samples were correctly classified through PCA-LDA analysis. In particular, we obtained 98% success rate for SK-MEL-2 cells and 90% for both SK-MEL-28 and MW-266-4 cells. Such outcome seems to reflect the spectral differences observed between the three cell types, which are starker in the comparison between *NRAS*-mutated cells and the *BRAF*-mutated ones. Moreover, the classification of both *BRAF* mutations is characterised by higher instability between the five rounds; this could be due either to “inadequacy” of the linear classifier in separating PC loadings, or to the biological variability between cells of the same type.

Overall, these findings show that SK-MEL-28 (bearing the classical *BRAF* V600E mutation), SK-MEL-2 (bearing the *NRAS* Q61R mutation) and MW-266-4 (bearing the rare *BRAF* V600D mutation) cells can be successfully identified through the information provided by their Raman signatures, that is, through the analysis of their molecular composition.

**FIGURE 2** A and B: scatter plots reporting PC loadings of each recorded spectrum (blue points = MW-266-4; green = SK-MEL-2; red = SK-MEL-28). The percentage of total variance explained by each PC is reported within round brackets. PC4 and PC11 loadings are reported in, A, while PC5 and PC8 are shown in, B. C, average auROCc value (black triangles) of each individual PC, and the percentage of the total variance of the dataset (magenta squares) explained by cumulating PCs from left to right. D, value of the NRMSE-CV for LDA (yellow diamonds) and ANN (orange circles) analyses, obtained by cumulating PCs from left to right



PCA-LDA	SK-MEL-2 (%)	SK-MEL-28 (%)	MW-266-4 (%)	ALL TYPES (%)
1st round	100.0	100.0	100.0	100.0
2nd round	90.0	70.0	90.0	83.3
3rd round	100.0	100.0	100.0	100.0
4th round	100.0	90.0	60.0	83.3
5th round	100.0	90.0	100.0	96.7
<b>Average</b>	<b>98.0</b>	<b>90.0</b>	<b>90.0</b>	<b>92.7</b>

Abbreviations: ANN, artificial neural network; PCA, principal component analysis; LDA, linear discriminant analysis.

*Note:* The dataset was divided among five subgroups, which were rearranged for five times in two sets – “learning” and “testing”. During each round, the learning set was used to train a PCA-LDA algorithm, which was later, applied to the classification of the corresponding testing set. The success rates in classifying SK-MEL-2, SK-MEL-28 and MW-266-4 samples are reported both separately and aggregated. The bottom row reports the average success rates of all five rounds.

### 3.4 | Cell type classification through PCA-ANN

Due to their capabilities in solving complex problems, ANNs are increasingly popular tools for the analysis of Raman datasets [19–21]. Hence, we applied them for discriminating MM cell lines, evaluating their classification accuracy and making a comparison with the results obtained through LDA.

The NNPR app automatically creates a three-layer network: an input layer for receiving the samples to be analysed; a hidden layer, whose number of neurons can be selected, for processing the input data; an output layer for generating a classification code. The NNPR tool iterates this process until the classification of all input data cannot be further improved. Moreover, the input dataset is divided into three categories:

- training – these samples train the network, which is adjusted according to the errors made in classifying them;
- validation – these samples are used to validate the network, and to halt training when their classification stops improving;
- testing – these samples have no effect on the network training and so provide an independent measure of its performance.

Finally, the NNPR routine compiles a classification algorithm based on the trained network, which can be used to analyse new, unknown input data.

During this analysis, we selected five neurons for the hidden layer. As previously stated, data were processed through PCA and NRMSE-CV estimation was used to select the nine PCs with more discriminatory power. Then, the whole dataset was organised in the exact same cross-validation rounds previously arranged for LDA.

**TABLE 1** Success rates of the PCA-LDA algorithms in classifying the samples belonging to the testing set (30 samples for each round of the 5-fold cross-validation)

During each round of the 5-fold cross-validation, the 120 samples of the learning set were given as input to the ANN and further partitioned between the three categories already mentioned (70% for training; 25% for validation; 5% for testing, since 0% is not available in the NNPR app). Then, the algorithm obtained from the ANN was used to classify the testing set (i.e., the remaining 30 samples). Table 2 reports the results of all five rounds. We achieved 96.7% success rate in classifying all the 150 samples; moreover, with respect to the previous PCA-LDA approach, we obtained a significant improvement – from 90% to 96% – in recognising classical *BRAF*-mutated cells (SK-MEL-28), while maintaining nearly the same accuracy for *NRAS*-mutated samples (SK-MEL-2, 96%). Such result is extremely important, because *BRAF* and *NRAS* mutations are involved in the vast majority of MM cases. A similar improvement – from 90% to 98% – was observed also in the success rate of MW-266-4 cells, providing better discrimination between different *BRAF* mutations. This is a relevant issue, since targeted therapy with *BRAF* inhibitors could be effective even in patients with non-V600E variations [22]. Thus, ANN algorithms allowed obtaining high classification accuracy between the three classes.

### 3.5 | Comparing LDA and ANN performances

Despite using very different approaches for processing data, LDA and ANN provided almost similar results: 94.7% (142/150) of the samples were classified – correctly or incorrectly – in the same way by both techniques. In particular, four out of 11 samples misclassified by LDA were misclassified by ANN, too, and the error was the same (i.e., the two algorithms attributed the same – incorrect – cell type to each sample). Moreover, the

**TABLE 2** Success rates of the PCA-ANN algorithms in classifying the samples belonging to the testing set (30 samples for each round of the 5-fold cross-validation)

PCA-ANN	SK-MEL-2 (%)	SK-MEL-28 (%)	MW-266-4 (%)	ALL TYPES (%)
1st round	100.0	100.0	100.0	100.0
2nd round	90.0	90.0	100.0	93.3
3rd round	100.0	100.0	100.0	100.0
4th round	100.0	90.0	90.0	93.3
5th round	90.0	100.0	100.0	96.7
<b>Average</b>	<b>96.0</b>	<b>96.0</b>	<b>98.0</b>	<b>96.7</b>

Abbreviations: ANN, artificial neural network; PCA, principal component analysis; LDA, linear discriminant analysis.

*Note:* The dataset was divided among the same five subgroups used in the PCA-LDA analysis, and rearranged for five times between learning and testing sets. The success rates in classifying SK-MEL-2, SK-MEL-28 and MW-266-4 samples are reported both separately and aggregated. The bottom row reports the average success rates of all five rounds.

remaining seven LDA errors correlated with deviations of the ANN scores from their target outputs: for example, the ANN score of one sample was (0.05; 0.52; 0.43) instead of the corresponding (0; 1; 0) target. These findings suggest that such misclassifications derived from real features in the Raman spectra, highlighted by both analytical techniques and possibly due to biological variability within the same cell line. More in general, we found that a linear classifier based on PCA is a simple but effective solution for discriminating MM cell types, while an algorithm based on ANN could help in improving the sensitivity to all lines (without “privileging” one over the other two).

### 3.6 | Discussion

In the last 10 years, Raman micro-spectroscopy has been applied for examining cells in a wide variety of biomedical studies, as it provides a fast, label-free and non-destructive molecular characterisation of the sample, which is alternative to both traditional (e.g. DNA sequencing, PCR) and new (e.g. mass spectrometry) analytical methods. This technique can be adapted to different experimental conditions: *in vivo* / *ex-vivo*, implemented in microscopes or cell sorting devices, using visible / near-infrared excitation wavelengths, etc. Moreover, the development of analytical tools based on machine learning allows unbiased, operator-independent and automated interpretation of the spectral information collected with such approach.

Within skin cancer research, several studies reported the successful discrimination of normal cells [10, 11] and tissues [23–25] from MM ones using Raman spectroscopy. Instead, we found only one work [10] addressing the discrimination of MM cell lines with different mutation profiles. Specifically, PCA loadings of different cell lines

were plotted into a 2D graph – as done here, in Figure 2 (A) and (B) – in order to show that their clusters could be spatially separated, but no attempt was made to build an algorithm for automated classification of cells through the analysis of their Raman spectra. By comparison, we used different instrumentation (a commercial microscope, instead of a custom-made one) and excitation wavelength (532 nm, instead of 784 nm), but the obtained Raman spectra and scatter plots are consistent with those reported in [10].

In this study, moreover, we provided a biochemical interpretation of the observed differences between different mutations, and the analysis of spectral data was further expanded in order to develop a classification algorithm for recognising MM profiles. Two alternative methods, based – respectively – on LDA and ANN, were implemented in MATLAB. We characterised their performances with different numbers of inputs/PCs and, then, compared their success rates in order to evaluate which analysis provided the highest classification accuracy. Overall, both methods reached high accuracy for all MM cell types (92.7% success rate for LDA, 96.7% for ANN). Based on the presented results, we found LDA to be a valid tool for this application; however, ANN seems better at diagnosing both classical (i.e., more frequent) and rarer *BRAF* mutations.

### 4 | CONCLUSION

This study suggests that Raman spectroscopy is capable of detecting molecular differences between melanoma cell types, and that it could provide a useful tool to guide the selection of the best molecular cell line model in research studies. Although the presented results need further validation (e.g. by increasing the number of cell lines per genotype), they show the possibility of optically

discriminating specific driver mutations in *BRAF* and *NRAS* genes, which is a very important clinical datum to be known for properly treating the disease.

## ACKNOWLEDGMENTS

This work was partially supported by the Italian Ministry for Education, University and Research in the framework of the Flagship Project NANOMAX, by the EU Horizon 2020 research and innovation programme under grant agreement No 654148 Laserlab-Europe, and by Ente Cassa di Risparmio di Firenze (private foundation).

## CONFLICTS OF INTEREST

The authors declare no financial or commercial conflict of interest.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## ORCID

Enrico Baria  <https://orcid.org/0000-0001-9679-7736>

## REFERENCES

- [1] American Cancer Society, Cancer Facts & Figures 2020, <https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2020.html>
- [2] A. Sandru, S. Voinea, E. Panaitescu, A. Blidaru, *J. Med. Life* **2014**, 7(4), 572.
- [3] G. C. Leonardi, L. Falzone, R. Salemi, A. Zanghi, D. A. Spandidos, J. A. McCubrey, S. Candido, M. Libra, *Int. J. Oncol.* **2018**, 52(4), 1071.
- [4] U. Banerji, A. Affolter, I. Judson, R. Marais, P. Workman, *Mol. Cancer Ther.* **2008**, 7(4), 737.
- [5] Skin Cancer Foundation, Melanoma treatment, <https://www.skincancer.org/skin-cancer-information/melanoma/melanoma-treatments> **2020**.
- [6] M. F. Berger, L. A. Garraway, *Hematol. Oncol. Clin. North Am.* **2009**, 23(3), 397.
- [7] T. Huang, J. Zhuge, W. W. Zhang, *Biomark. Res.* **2013**, 1, 3.
- [8] R. Smith, K. L. Wright, L. Ashton, *Analyst* **2016**, 141, 3590.
- [9] K. Hamada, K. Fujita, N. Smith, M. Kobayashi, Y. Inouye, S. Kawata, *J. Biomed. Opt.* **2008**, 13(4), 044027.
- [10] E. Brauchle, S. Noor, E. Holtorf, C. Garbe, K. Schenke-Layland, C. Busch, *Clin. Exp. Dermatol.* **2014**, 39(5), 636.
- [11] H. Wang, T. H. Tsai, J. Zhao, A. M. Lee, B. K. Lo, M. Yu, H. Lui, D. I. McLean, H. Zeng, *Photodermatol. Photoimmunol. Photomed.* **2012**, 28(3), 147.
- [12] C. Krafft, C. Beleites, I. W. Schie, J. H. Clement, J. Popp, *Proc. SPIE* **2016**, 9704, 970408. [https://spie.org/Publications/Proceedings/Paper/10.1117/12.2217781?origin\\_id=x4323&start\\_year=1963&SSO=1](https://spie.org/Publications/Proceedings/Paper/10.1117/12.2217781?origin_id=x4323&start_year=1963&SSO=1)
- [13] M. C. Sini, V. Doneddu, P. Paliogiannis, M. Casula, M. Colombino, A. Manca, G. Botti, P. A. Ascierto, A. Lissia, A. Cossu, G. Palmieri, *Oncotarget* **2018**, 9(9), 8531.
- [14] D. Tuschel, *Spectroscopy* **2016**, 31(3), 14.
- [15] J. Zhao, H. Lui, D. I. McLean, H. Zeng, *Appl. Spectrosc.* **2007**, 61(11), 1225.
- [16] R. Bro, A. K. Smilde, *Anal. Methods* **2014**, 6, 2812.
- [17] W. Liu, Z. Sun, J. Chen, C. Jing, *J. Spectrosc.* **2016**, 1603609. <https://www.hindawi.com/journals/jspec/2016/1603609/>
- [18] M. Gniadecka, P. A. Philipsen, S. Sigurdsson, S. Wessel, O. F. Nielsen, D. H. Christensen, J. Hercogova, K. Rossen, H. K. Thomsen, R. Gniadecki, L. K. Hansen, H. C. Wulf, *J. Invest. Dermatol.* **2004**, 122(2), 443.
- [19] J. Liu, M. Osadchy, L. Ashton, M. Foster, C. J. Solomon, S. J. Gibson, *Analyst* **2017**, 142(21), 4067.
- [20] M. Fukuhara, K. Fujiwara, Y. Maruyama, H. Itoh, *Anal. Chim. Acta* **2019**, 1087, 11.
- [21] W. Lee, A. T. M. Lenferink, C. Otto, H. L. Offerhaus, *J. Raman Spectrosc.* **2019**, 51(2), 293.
- [22] R. Cicchi, A. Cosci, S. Rossari, D. Kapsokalyvas, E. Baria, V. Maio, D. Massi, V. De Giorgi, N. Pimpinelli, F. S. Pavone, *J. Biophotonics* **2014**, 7, 86.
- [23] L. Heinzerling, S. Kühnapfel, D. Meckbach, M. Baiter, E. Kaempgen, P. Keikavoussi, G. Schuler, A. Agaimy, J. Bauer, A. Hartmann, F. Kiewewetter, R. Schneider-Stock, *Br. J. Cancer* **2013**, 108(10), 2164.
- [24] C. Yorucu, K. Lau, S. Mittar, N. H. Green, A. Raza, I. Rehman, S. MacNeil, *Appl. Spectrosc. Rev.* **2016**, 51(4), 243.
- [25] J. Zhang, Y. Fan, Y. Song, J. Xu, *Medicine (Baltimore)* **2018**, 97(34), e12022.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Baria E, Cicchi R, Malentacchi F, et al. Supervised learning methods for the recognition of melanoma cell lines through the analysis of their Raman spectra. *J. Biophotonics*. 2021;14:202000365. <https://doi.org/10.1002/jbio.202000365>