scientific reports

Check for updates

OPEN

Machine learning to detect melanoma exploiting nuclei morphology and Spatial organization

Giulia Veronesi[1,2,6], Nico Curti[3,6], Aldo Gardini[4]✉, Giulia Querzoli[5], Gastone Castellani[1,6] & Emi Dika[1,2,6]

Cutaneous melanoma is one of the most lethal forms of skin cancer, and its incidence is increasing globally. Its diagnosis typically relies on manual histopathological examination, a process that is both complex and time consuming. In this study, we propose an automated diagnostic tool, capable of generating interpretable results to aid clinical decision-making. A total of 146 whole slide images are included in the study, encompassing various lesion types: congenital nevi, dysplastic nevi, melanomas, and melanomas on nevi. The images were first processed using a multi-resolution image processing pipeline with the aim of segmenting nuclei, evaluating their geometrical and morphological features, as well as their spatial organization. To characterize each slide, these features were synthesized into 44 variables, which were then subjected to Linear Discriminant Analysis. Through this procedure, 18 relevant variables were identified demonstrating good performance in melanoma detection, as validated through Monte Carlo Cross-Validation. These variables were also interpreted within the framework of established histopathological diagnostic insights. By refining the analysis to the cellular level, we emulated standard clinical evaluation practices, ensuring that every aspect of the diagnostic process was accessible and verifiable by medical professionals. The proposed tool can offers significant potential to support clinicians in various tasks, such as prioritizing the analysis of critical samples and providing a secondary diagnostic opinion in complex cases.

Keywords Melanoma, Machine aided diagnosis, Machine learning, Spatial statistics, Histopathology

The incidence of Cutaneous Melanoma (CM) has been steadily rising globally over the past few decades[1–4]. The current gold standard for melanoma diagnosis is the histopathological examination of hematoxylin and eosin (H&E) stained tissue biopsy sections under light microscopy. This method allows pathologists to observe the disease at the cellular level, leveraging their expertise to assess the tissue's morphological and cytological features for accurate diagnostic assessment[5].

However, manual evaluation of tissue samples is often complex, making it both time-intensive and labor-intensive process. Additionally, the high volume of skin biopsies processed in most pathology laboratories poses substantial logistical and workforce challenges. Interpretation and classification rely on the integration of numerous histological features[6], a process that can be subject to pathologist subjectivity. As a result, diagnostic discrepancies in melanocytic neoplasms have been reported to range from 2.7% to as high as 25–26%[7–9]. The significant intra- and inter-observer variability in the evaluation of melanocytic skin lesions can lead to both underdiagnosis and overdiagnosis of melanoma[10–12].

Observer variability is most commonly encountered in the diagnosis of small or flat lesions[8,12–16]. There is a tendency to over-diagnose thin melanomas due to considerable challenges in differentiating junctional-type dysplastic nevi and in situ/superficial spreading melanoma[17,18]. Further diagnostic complexity is posed by melanomas that are difficult to diagnose, such as spitzoid melanoma or melanoma arising from a nevus.

[1]Department of Medical and Surgical Sciences, University of Bologna, Bologna 40126, Italy. [2]Oncologic Dermatology Unit, IRCCS, Azienda Ospedaliero Universitaria Bologna, Bologna 40138, Italy. [3]Department of Physics and Astronomy, University of Bologna, Bologna 40127, Italy. [4]Department of Statistical Sciences, University of Bologna, Bologna 40126, Italy. [5]Pathology Unit, IRCCS Azienda Ospedaliero-Universitaria di Bologna, Bologna 40126, Italy. [6]Giulia Veronesi, Nico Curti, Gastone Castellani and Emi Dika contributed equally. ✉email: aldo.gardini@unibo.it

In this context, the development of new diagnostic techniques is crucial to assist pathologists in their daily practice. The advent of digital pathology and the implementation of whole slide images (WSI)[19], have paved the way for the application of artificial intelligence to histopathology, i.e. the so-called Pathomics, which holds considerable promise in addressing these issues, by providing an objective support to clinicians[20–23].

Several studies have demonstrated the potential of automated tools as clinical decision support systems, particularly in the automated detection and classification of cells and tissues. While numerous AI-based applications have been developed for various cancer types[24,25], the field of dermatohistopathology remains relatively underexplored[26]. Only a limited number of studies have progressed to clinical use, primarily due to the lack of generalizability and transparency in the methodological employed. A major challenge remains in replicating the complex deductive reasoning of expert pathologists, which involves extracting, integrating, and interpreting morphological and spatial features in a statistically robust and clinically meaningful way. In contrast, current trends in Pathomics research often prioritize classification performance using deep learning neural networks[8,27–30], with interpretability addressed only retrospectively or as a secondary goal[27,31–33].

The lack of interpretability in AI models poses a significant barrier to their clinical adoption. Many of these models operate as "black boxes", generating decisions without providing clear explanations that align with clinical reasoning or histopathological logic[33,34]. For example, large-scale studies in skin cancer have demonstrated the potential of AI for melanoma classification[29,31], with promising results in terms of diagnostic performance and reproducibility. However, such models often rely on patch-based training, which results in the loss of fine-grained, cell-level resolution. Consequently, the outputs are restricted to regional predictions rather than detailed cellular analyses. In some cases, efforts have been made to improve explainability — such as through algorithmic strategies based on traditional machine learning or rule-based systems[28] — but these often require extensive manual steps, limiting their scalability.

Another critical limitation of current AI models is their insufficient capacity to integrate and leverage spatial information within the tumor microenvironment. In particular, the spatial distribution of lymphocytes in relation to tumor cells — including melanocytes — and the surrounding stromal or necrotic regions holds significant prognostic value. Evidence from other tumor types suggests that immune cell localization within specific tissue compartments, rather than their mere presence, is predictive of treatment response.

Despite the growing interest in AI applications within dermatopathology, the lack of publicly available, well-annotated datasets[35] and the limited sharing of trained models represent significant obstacles to the reproducibility and validation of proposed methods. Many existing studies rely on proprietary data or use small cohorts with limited diversity in terms of tissue types, staining protocols, or diagnostic categories[36–43]. Furthermore, access to source code and pretrained models is often restricted or available for only no-melanoma specific architectures[44,45], preventing the scientific community from performing independent evaluations or benchmarking new algorithms against established baselines. This lack of openness hinders transparency, slows methodological progress, and limits the translation of promising research into clinical practice. Establishing standardized protocols for data sharing, model reporting, and interpretability evaluation is therefore essential to ensure the development of robust, generalizable, and clinically trustworthy AI tools in dermato-histopathology.

In the context of melanoma, a highly aggressive skin cancer with increasing incidence[46], AI offers a valuable opportunity to enhance the assessment of key prognostic markers such as tumor-infiltrating lymphocytes (TILs). TILs have been associated with both improved response to immune checkpoint inhibitors and longer survival times in metastatic melanoma[47–52], as well as reduced recurrence in primary melanoma[53]. Yet, their current clinical evaluation is typically manual, based on qualitative scoring systems such as the Clark scale[51,54,55], which are subject to substantial interobserver variability[51,56,57]. Deep learning models may provide more consistent and detailed assessments; however, several limitations persist. Chief among these is the inability of existing models to integrate spatial context—specifically, the localization of TILs and their proximity to tumor cells such as melanocytes.

The spatial distribution of lymphocytes in relation to melanocytes, tumor nests, stromal regions, or necrotic areas plays a critical role in prognostication, as demonstrated in other cancer types such as breast and lung cancer[39,40]. In melanoma, the intra-tumoral versus peritumoral positioning of immune cells, as well as their interactions with melanocytes, could yield essential diagnostic and therapeutic insights. Furthermore, subsets of immune cells, such as neutrophils and B cells, have shown relevance in predicting metastatic potential and response to therapy[41,42]. Despite these observations, current AI models often fail to deliver such spatially resolved insights, largely due to the absence of annotated datasets and algorithms capable of jointly segmenting nuclei and tissue regions with high fidelity. The segmentation of both cellular components and their surrounding microenvironments is thus a prerequisite for meaningful and interpretable AI applications in clinical practice.

In this study, we propose a novel, fully automated analysis pipeline that integrates artificial intelligence and statistical methods to prioritize the review of dermatopathology slides. The goal is not only to improve melanoma classification accuracy but also to ensure full interpretability of the variables and results. The pipeline is entirely open-source and emulates clinical evaluation practices by refining the analysis down to the cellular level. In doing so, it captures critical morphological and spatial features of cell populations — particularly melanocytes and lymphocytes — while maintaining transparency in every step of the diagnostic process. The proposed methodology aims to assist clinicians in identifying high-risk histological features and distinguishing malignant lesions from benign nevi, with potential applications in complex differential diagnoses, such as dysplastic nevi versus cutaneous melanoma.

In the following sections we will describe the analyzed dataset, summarizing the key characteristics of the considered population and outlining the specifications of the hardware employed for WSIs digitization. Each slice was analyzed using an advanced image processing pipeline designed to automatically identify and segment all nuclei present in the tissue using a custom convolutional neural network architecture. For each identified nucleus, a series of morphological and textual features were quantified, allowing a high-resolution description

of the sample. The extracted features were further summarized using statistical indicators, providing a robust and interpretable overview of the spatial organization of nuclei in the tissue. The effectiveness of these indicators in distinguishing dermatological categories was evaluated by incorporating them into a classification model. In conclusion, we discuss the limitations and the main contributions of our findings, along with their clinical implications.

Materials and methods

Study population and patient selection

Given the retrospective nature of the study, the Institutional Review Board of the IRCCS Sant'Orsola-Malpighi Hospital, University of Bologna (Italy), waived the requirement for informed consent. The study and data acquisition protocol were approved by the Local Ethics Committee (Comitato Etico Area Vasta Emilia Centro, CE-AVEC), and conducted in accordance with the Declaration of Helsinki (protocol reference identifier 'Skincancer2019'). The study involved a population of 100 patients (53 males and 47 females, with a mean age of 58.3 and 52 years, respectively). Histological glass slides of 20 Congenital Nevi (CN), 40 Dysplastic Nevi (DN), 30 CM, and 10 CM on Nevus (CM +N) were collected and digitized using a Nano-Zoomer 2.0-RS Hamamatsu scanner with a 40 ×(0.23 μm/pixel) magnification and autofocusing. H&E-stained WSIs were reviewed independently by two expert histopathologists, each possessing more than 10 years of experience in dermato-oncology.

Criteria for inclusion were the following: (i) Patients diagnosed between January 2022 and December 2023 with available histopathological slides in good condition; (ii) CM with superficial spreading histological type; (iii) Lesions located in conventional body sites (excluding mucosal, acral, and facial areas).

To ensure a comprehensive evaluation for clinical practice, multiple WSI slices were collected for each patient. To reduce data redundancy, only non-consecutive slices were considered for the statistical analysis of each patient. This approach allowed us to increase the number of available samples for each class, working under the assumption that the spatial structure of the lesions remains sufficiently distinct when sampled from different biopsy regions. Using this approach, a total of 146 WSI samples were collected for analysis.

Histological image analysis

The complete set of WSIs was analyzed using a multi-resolution image processing pipeline. Starting from the low-resolution level of the WSI pyramid, the entire set of slices (region-of-interests, ROIs) was automatically identified using an Otsu thresholding algorithm[58] applied to the grayscale version of each image. Small connected components, corresponding to spurious artifacts or debris on the glass slide, were filtered out. The identified slice ROIs were used to extract image information at the highest resolution level (40× magnitude), thereby enabling subsequent analysis to focus only on the informative regions of the WSI. For each slice, a binary mask of the entire tissue was extracted, providing a coarse grain representation of the slice contours. Given that all the images originated from a single center (using the same laboratory and scanner for digitization) and given the high quality of the data, no additional data pre-processing steps were required.

According to the slice mask obtained at the low-resolution level of the WSI, a dedicated U-Net segmentation model[59] was applied to identify individual cells at the highest resolution level. The model was pre-trained on the PanNuke[45] dataset, which consists of more than 7.5k images and their corresponding masks. It employs a loss function specifically designed to minimize misleading segmentations caused by to overlapping components (the IoU score achieved on 20% of the images used as validation set is 0.81 ±0.09). This approach aims to achieve a detailed delineation of cellular contours, while avoiding possible unrealistic shapes. The model was trained on 256 ×256 RGB images and applied to all identified ROIs: each ROI was split into a series of overlapping patches (stride 128 ×128), generating a probability map for nuclei segmentation that was converted into binary values via thresholding at 0.9: on overlapping areas, the maximum probability value was retained as the reference for the thresholding procedure. A detailed description of the model and the implemented loss function falls beyond the scope of this work, and will be further discussed in a future publication. To ensure accuracy, masks obtained from the model were randomly reviewed by expert histopathologists, confirming their concordance with manual annotations.

The entire image processing pipeline was developed in Python (v3.10.4) using Openslide[60] (v3.4.1, accessed on 16 January 2025) and OpenCV[61] (v4.5.5, accessed on 16 January 2025) packages for the management and analysis of WSI samples. The entire segmentation pipeline was performed using a 64-bit workstation machine (64 GB RAM memory and 1 CPU i9-10920X Intel®, with 12 cores, and a Quadro RTX 4000 NVIDIA® GPU): under this setup, the computational time required for processing an entire slice was around 5 min.

Nuclei feature extraction

We applied the nuclei segmentation model to the entire dataset, collecting more than 6 million segmentations. Each segmented component, i.e. a nucleus, was analyzed by characterizing its morphological and textural aspects (ref. Figure 1). For each nucleus, the corresponding contour of the binary mask was extracted. The component was then characterized based on 16 morphological features (e.g. area, perimeter, circularity, elongation, etc.) and 37 textural features (e.g. RGB, HSV statistics, and Haralick features[62]. The detailed list of features is available in the *Supplementary Material* of this manuscript. Due to the high variability in slice staining (e.g., color contrast, staining degradation, tissue stain capture, etc.), only the morphological features were used in the next analysis.

The spatial localization of each nucleus was estimated based on the barycenter of its contour shape. Applying the Principal Component Analysis (PCA)[63] to the entire set of nucleus coordinates, each slice was rotated to align the epidermis surface with the top of the image. Slice orientation plays a crucial role in dermato-histological analysis, facilitating the evaluation of nucleus proximity to the epidermis layer, and the use of PCA is based

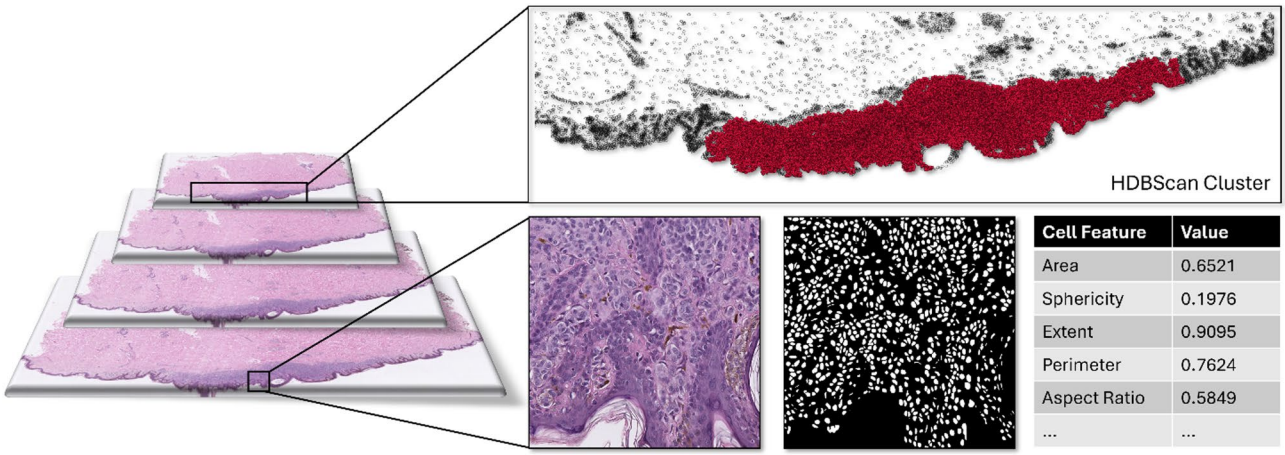

Fig. 1. Histological Image Analysis Pipeline. *Schematic representation of the histological image analysis pipeline developed for lesion identification. The analysis is performed on the multi-resolution WSI, starting from the identification, segmentation, and characterization of the entire set of nuclei at the high-resolution level of the image pyramid. Each nucleus is characterized by a series of morphological and textural features. The spatial coordinates of each nucleus are used to identify the lesion area via clustering method (HDBScan), aiming to filter the lesion ROI and its silhouette at the low-resolution level of the image pyramid.*

on the assumption that the majority of the nuclei in each slice are distributed in the upper layers of the tissue. The spatial coordinates of each nucleus were subsequently rescaled according to the PCA coordinate system, providing information about their spatial distribution in relation to the epidermis layer.

Describing samples: variables computation

The output of the nucleus segmentation process was further synthesized to build a set of variables describing the peculiarities of the different slides. This statistical processing of the samples, along with the subsequent analyses, was conducted using R (v4.4.1 running under Windows 11 ×64 machine). Firstly, the lesion in each slice was identified as the larger clusters of nuclei detected using the HDBScan algorithm[64]. This method is a density-based clustering algorithm that effectively identifies groups of units (nuclei, in this case) that are connected based on density related measures.

Subsequently, focusing on the nuclei constituting the lesion, the behaviors of nuclei area (A), aspect ratio (AR), extent (E), and sphericity (S) were investigated, noting that their spatial distributions vary across the sample area (see Fig. 2a, for an example on area). Given that these studied lesions are known to exhibit different nucleus characteristics with respect to their depth, i.e. their distance from the top, we analyzed them from this perspective. As illustrated in Fig. 2b (blue lines), non-linear trends emerge in the relationships between the considered features and the distance of nuclei from the surface. To extract valuable information for statistical learning models, these patterns were summarized by splitting the nuclei into different strata, namely four groups (labeled with *S1*, *S2*, *S3*, and *S4*) based on by quartiles of the distances (Fig. 2c). Due to the irregular behavior observed in the fourth group of nuclei, it was excluded from the analysis. Within each stratum, the mean and standard deviation (*sd*) of nucleus attributes were computed (e.g. *E_S2_mean* indicates the mean nucleus extent in the second stratum). To assess the heterogeneity of these features across the strata, the between-strata standard deviation was also computed (named, e.g., *E_sd_bet*), defined as the standard deviation of the means obtained across the different levels of nuclei.

A further useful aspect to include in the analysis concerns the spatial correlation of the features measured on the segmented nuclei, i.e. whether similar nuclei tend to be spatially clustered. To this aim, basic descriptive geostatistical tools[65] were used and we fitted the Spherical and the Exponential variogram models on the empirical variogram using the R package *gstat*[66] (v2.1). The best variogram model was selected based on the Akaike Information Criterion. The area feature showed a marked spatial correlation that we summarized in three typical variogram parameters: *A_Range*, *A_Nugget*, and *A_Sill*. To further describe the relationship among nuclei, we computed additional variables based on the distribution of the nearest neighbor distances (NND), evaluated using the *spatstat* R package[67] (v3.1). The NND distributions were summarized in the whole lesion and in the *central* part (S2 + S3) using the median (*med*), standard deviation, and first and third quantiles (*q1* and *q3*). Lastly, we incorporated the number of nuclei, the strata widths, and the density of nuclei in S1. As a result, 44 variables were identified, each describing the lesions within the sample.

Classification model

We studied several supervised classification techniques (see *Supplementary Materials* for further details), and determined that the most interesting results were obtained using the set of methods that includes Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA)[68,69]. In more detail, we applied LDA to the entire set of variables, obtaining an optimal subset through a stepwise forward selection algorithm based

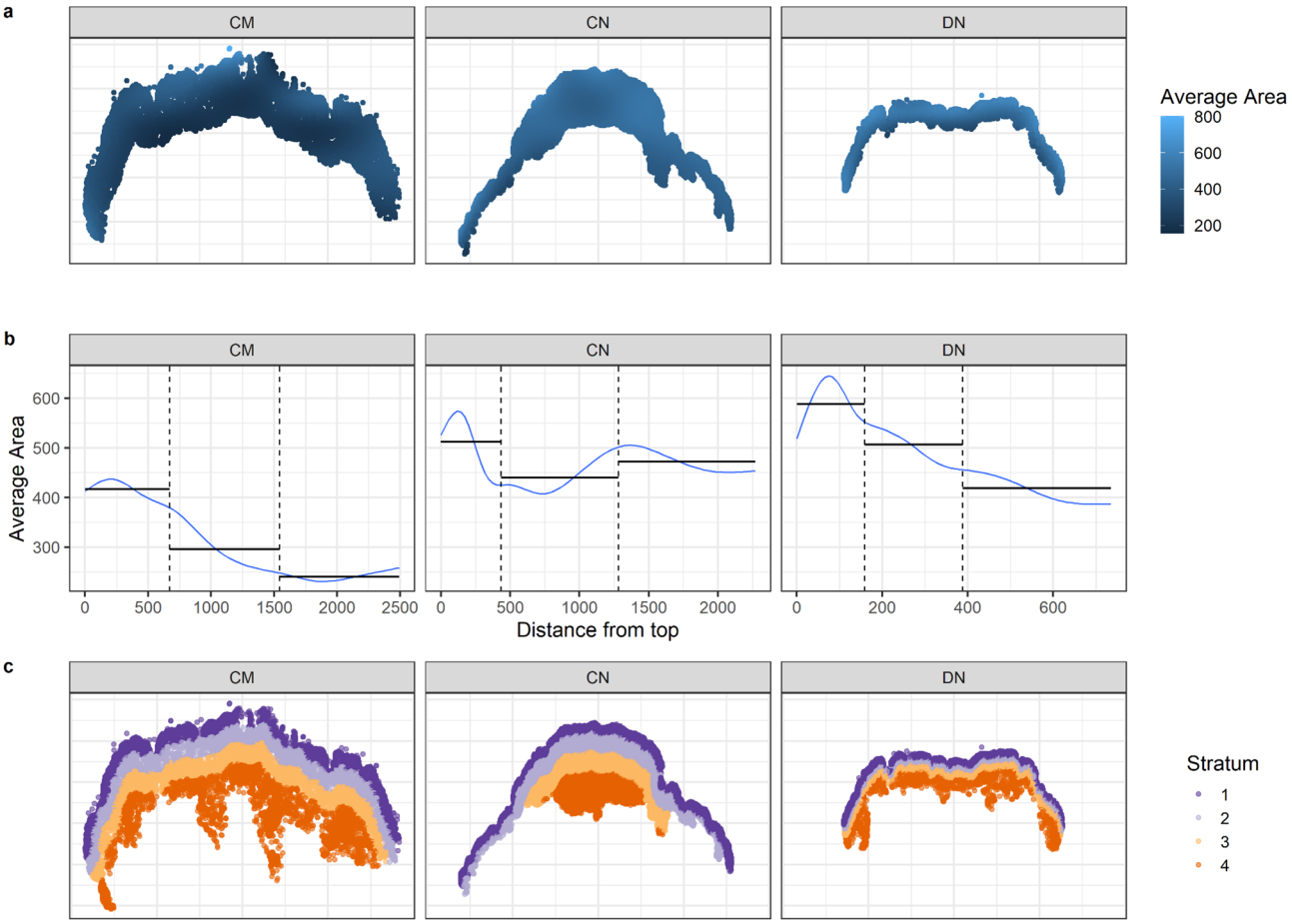

Fig. 2. Computing Variables for Samples Profiling. *Graphical representation of the steps used to explain how the variables describing the samples were computed. The area of nuclei was used as an example, with each plot shown for three slides: a CM, a CN, and a DN.* (**a**) *Kernel density plot of the average area distribution of nuclei across the slides.* (**b**) *Relationship between the average area of nuclei and their distance from the top: the blue line represents the cubic spline estimate, and the black horizontal lines indicate the averages of nuclei grouped into strata S1, S2, and S3.* (**c**) *Nuclei grouped into strata (S1, S2, S3, and S4), based on quartiles of their distance from the top.*

on Wilks's Lambda criterion. As a benchmark, we also considered Regularized Discriminant Analysis (RDA)[70], which automatically mitigates potential multicollinearity issues and combines the group covariance matrices used in LDA with the broader definition of QDA. However, this comes at the cost of reduced interpretability of the discriminant directions that characterize LDA. These classification methods provide a vector of probabilities for each sample which is then assigned to the most probable class using a Bayes classifier approach. To implement these methods we used functions available in R within the packages MASS[71] (v7.3) and klaR[72] (v1.7).

To assess the effectiveness of classification methods, a Monte Carlo Cross-Validation (MCCV) study was performed by randomly splitting the whole dataset 200 times into a training set (85%) and a test set (15%). This choice was primarily influenced by the relatively small sample size. The trained models were then applied to classify the samples in the test set, generating a confusion matrix comprising false negatives (FN), false positives (FP), true negatives (TN), and true positives (TP). At each iteration, we computed the following metrics:

$$accuracy = \frac{TN + TP}{FN + FP + TN + TP},$$

$$sensitivity = \frac{TP}{TP + FN},$$

$$precision = \frac{TP}{TP + FP},$$

$$F_1 = \frac{2TP}{2\,TP + FP + FN},$$

Variable label	Variable description	Group means		
		CM	DN	CN
A_S1_mean	Average area in S1	−0.107[a]	0.246[b]	−0.419[b]
A_S2_sd	Standard deviation of area in S2	0.081[a]	0.220[a]	−0.684[b]
A_S3_sd	Standard deviation of area in S3	−0.331[a]	0.529[b]	−0.723[c]
A_S3_mean	Average area in S3	−0.275[a]	−0.080[b]	0.678[b]
A_sd_bet	Standard deviation of area between strata	−0.031[a]	0.450[b]	−1.055[c]
AR_S2_sd	Standard deviation of aspect ratio in S2	0.260[a]	−0.158[b]	−0.065[b]
S_S2_sd	Standard deviation of sphericity in S2	0.242[a]	−0.155[ab]	−0.041[b]
S_S2_mean	Average sphericity in S2	−0.321[a]	0.139[a]	0.218[b]
S_S3_mean	Average sphericity in S3	0.229[a]	−0.355[a]	0.474[b]
E_S3_sd	Standard deviation of extent in S3	−0.334[a]	0.445[a]	−0.512[b]
E_sd_bet	Standard deviation of extent between strata	−0.366[a]	0.566[b]	−0.754[c]
A_Range	Range parameter of the variogram related to area	0.498[a]	−0.338[b]	−0.040[c]
A_Nugget	Nugget parameter of the variogram related to area	−0.037[a]	0.338[a]	−0.769[b]
Width_S1	Width of S1	0.263[a]	−0.464[a]	0.684[b]
Width_S2	Width of S2	0.158[a]	−0.372[a]	0.640[b]
NND_central_med	Median of NND, considering nuclei in S2 and S3	0.100[a]	0.087[ab]	−0.390[b]
NND_central_q1	First quartile of NND, considering nuclei in S2 and S3	−0.100[a]	0.124[a]	−0.130[a]
NND_q3	Third quartile of NND, considering all nuclei	0.275[a]	0.095[b]	−0.716[c]

Table 1. Variables relevant for LDA. *Labels and descriptions of the 18 relevant variables selected for LDA using a Stepwise algorithm. For each variable, the average value across the different types of lesions (CM, CN, and DN) is provided, along with a letter indicating groups that are significantly different based on a Kruskal-Wallis test.*

	Accuracy	Sensitivity	Precision	MCC	F1 Score
LDA	0.904	0.844	0.865	0.783	0.842
RDA	0.877	0.850	0.796	0.730	0.808

Table 2. Performances of classification methods. *Average values of the performance indicators evaluated in the Monte Carlo Cross-Validation study.*

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}},$$

and we reported their average as output.

Results

We applied the proposed image processing pipeline to a dataset consisting of 146 histological slices, split into 28 CN, 69 DN, 35 CM, and 14 CM + N, analyzing the spatial organization of the nuclear morphological features. The first step of the analysis involved the identification of tissue areas at the low-resolution level of the WSI pyramid. Starting from the identified ROIs, a deep learning U-Net architecture was implemented to segment all nuclei within the tissue slices, applying the described methodology to compute the summary variables.

Classification

To evaluate the effectiveness of the set of computed variables, we applied classification methods to categorize samples into three groups: CM (including also CM + N), DN, and CN. LDA was performed on the data with two goals: achieving satisfactory classification performance and understanding how the computed variables characterize the groups. The 44 covariates were standardized (linearly transformed to obtain zero mean and unitary variance) before running the analysis. A preliminary variable selection step identified 18 relevant variables, listed in Table 1.

To assess the effectiveness of the classification methods (CM vs. non-CM), we conducted the MCCV study, whit results presented in Table 2. By labeling melanoma samples as 'positive', the sensitivity index (0.844) provided insight into the effectiveness of the classification method in detecting actual cases. This high sensitivity was complemented by a precision of 0.865, indicating that the classification method did not substantially overestimate the number of predicted melanomas. As a comprehensive measure of case detection, the obtained F1 score was good (0.842). Additionally, the overall classification accuracy was high (0.904), further supported by a robust MCC of 0.783. We also highlight that LDA, when performed on an optimal subset of variables outperforms RDA based on the evaluated metrics. To further validate the classification algorithm, we compared

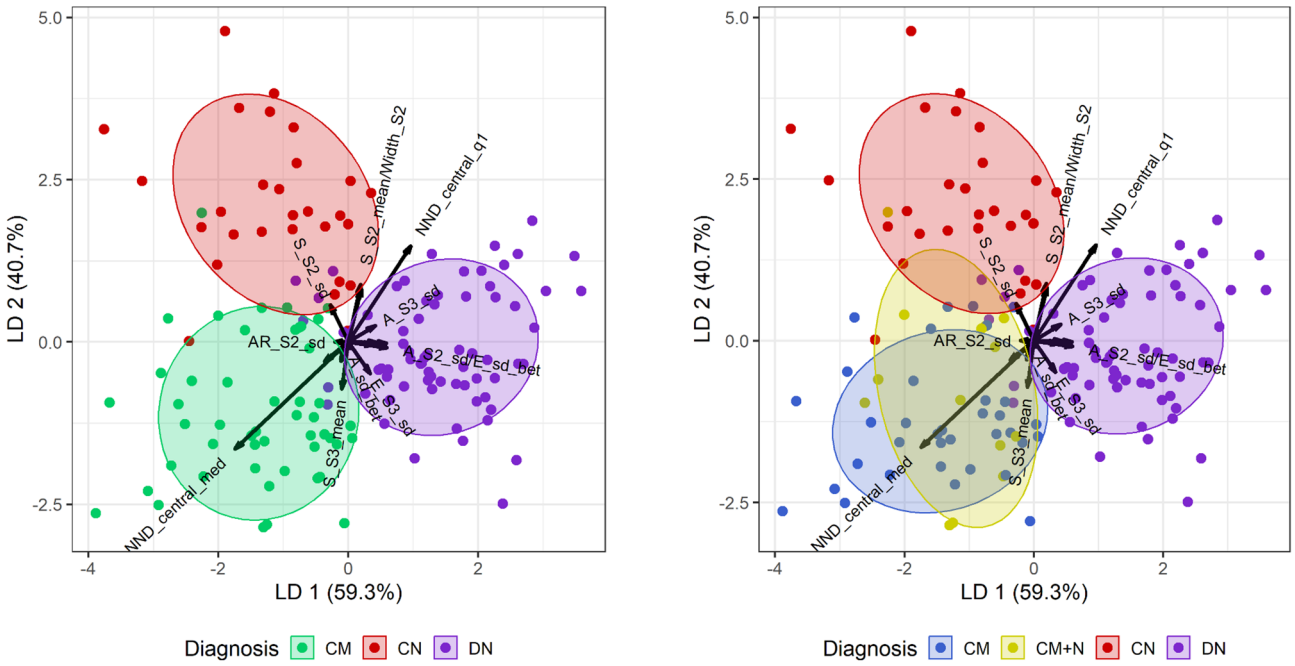

Fig. 3. LDA bi-plot. *Scores of the samples projected in the 2-dimensional space produced by the LDA, reporting the 95% confidence ellipses to highlight the groups. Arrows indicate the direction of the most relevant variables for the samples discrimination (discriminant coefficients).*

these results with those obtained with other widely used models, including Random Forest, Boosting, and Support Vector Machines. The results, available in the *Supplementary Materials* along with the hyperparameter settings, further confirmed the strong performance of LDA.

A key advantage of LDA is its interpretability, as it leverages dimensionality reduction properties of the technique to shed light on the classification process. Figure 3 shows the scores of the entire dataset projected onto the two-dimensional space defined by LDA, which represents linear combinations of the original variables that maximize the ratio between-group to within-group variance. The 95% confidence ellipses indicated effective group separation, with the right-hand-side plot revealing CMs near the CN group, suggesting the presence of CM cases overlapping with CN. The first linear discriminant (LD) axis accounts for 59.3% of the between-group to within-group variance ratio, mainly separating DN (on the right) from CN and CM (on the left), with CN showing higher scores along the second LD axis. Figure 3 also shows the discriminant coefficients (arrows), which indicate the directions of the most relevant variable contributions along the LD axes. This output should be interpreted alongside Table 1, which lists the mean values of variables for each sample group. Superscript letters in the table denote groups that are significantly different, based on post-hoc comparisons using the Kruskal-Wallis test ($a=0.05$), with the p-values adjusted for multiple comparisons via the multivariate t method.

Discussion

The emergence of digital pathology has enabled the development of increasingly efficient automated models that enhance, accelerate, and standardize the diagnostic procedure. In onco-dermatology, critical challenges such as distinguishing CM from benign lesions (nevi) are addressed with increasingly high accuracy, thanks to the rise of artificial intelligence[27,33]. However, this trend towards more complex and high-performance models often conflicts with current demands from the medical-scientific community, which requires complete transparency in decision-making processes[73,74].

The field of automated analysis is dominated by models based on convolutional neural networks and deep learning architectures, both of which are known for their limitations in explainability. Efforts to improve the transparency of these models often require significant resources and, in many cases, are only addressed after performance metrics have been quantified, with the hope of retroactively justifying prior decisions[28,31].

Another significant challenge associated with neural network-based models is the substantial amount of data required for training. To address this issue, constant effort is demanded from expert clinicians to annotate, segment, and evaluate digitized images[29]. The use of public and shared databases aims to mitigate this demand, but the quality of available data is often insufficient for some certain applications, and inter-clinician discordance limits the models' generalization capabilities[12]. The data annotation challenge is often bypassed using classification models that consider portions of the image (patches), producing a tiling of the space[8,29,31] or through pixel clustering techniques on the image[28]. Despite the excellent results achieved with these approaches, the main drawback remains the coarse quality of the generated descriptions. The lack of a pixel-wise and high-resolution characterization of the tissue limits the quality of extracted information, undermining the necessary link between human and artificial deductive processes.

In this work, our goal was to enhance the interpretability of variables extracted for both nuclear morphology and spatial organization. The significance of spatial information had already been demonstrated in our previous research[23], aligning with the clinical diagnostic process. By refining our analysis down to the cellular level, we were able to develop a method that closely mirrors clinical evaluation practice. This workflow ensures clinical validation at each stage of analysis but also maximizes flexibility, allowing human intervention whenever inconsistencies occur.

The classification results were obtained using a Bayes classifier approach to maximize overall accuracy. However, from a clinical perspective, this approach may not be ideal, as the two types of classification errors have different levels of severity. For instance, failing to diagnose a case of CM is far more critical than misclassifying a benign lesion as CM. To address this imbalance, the threshold for assigning the CM label can be lowered to increase sensitivity. However, this comes at the cost of reducing precision and, subsequently, lowering overall accuracy.

The main limitation of our work could be imputed by the monocentric nature of the dataset and to the limited heterogeneity of the selected patients. Therefore, a more extensive validation of the proposed pipeline is mandatory before its possible integration into a clinical workflow. Directly related to these requirements, there is a second limitation of this study due to the use of a single scanner, which may introduce biases in image acquisition protocols. To improve the robustness of the system, validation should include samples scanned in different conditions and with greater hardware heterogeneity. A similar consideration could be made taking into consideration the moderately limited number of samples considered, which could affect the generalizability of our pipeline. Another challenge could be imputed to the inability to achieve fully accurate validation of the segmentations. To mitigate this source of errors we employed a statistical approach, but further testing is necessary to quantitatively assess the validity of our assumptions. All identified limitations will be addressed in future studies, improving the image processing pipeline and enlarging the dataset with new records according to the clinical availability.

An important outcome of the implemented methodology is the ability to interpret the classification rule from a histopathological perspective, as LDA is not considered a "black-box" approach. In this study, a histopathologist validated the diagnostic significance of melanoma based on the morphological features and spatial distribution of nuclei. For instance, one of the most critical aspects of melanoma diagnosis is the nuclear pleomorphism of atypical melanocytes. The nuclei tend to have an irregular shape, sometimes with the presence of nucleolus and thickening of the nuclear membrane. The greater variability in nuclear area and sphericity in the second skin layer (*A_S2_sd* and *S_S2_sd*) can be attributed to increased pleomorphism observed in CM or DN, while CN is characterized by more uniform nuclear shapes. This observation aligns with histopathological findings, where increased nuclear irregularity and atypical melanocytes are strong indicators of malignancy. Another key feature analyzed was the spatial distribution of nuclei and the distance between them. The median nearest-neighbor distance (*NND_central_med*) showed that melanoma nuclei exhibit a greater degree of spatial separation supporting histopathological findings, where loss of cellular cohesion is recognized as a significant marker of malignant transformation. These results emphasize the importance of cellular and nuclear composition in distinguishing between benign and malignant skin lesions in dermatopathology. Artificial intelligence in dermatopathology cannot replace the human element in medical care. However, the integration of automated machine learning in histopathology combined with the development of novel biomarkers in laboratory research, can provide clinicians with a valuable tool for improving precision medicine and patient management[75–78].

Automated clinical procedures provide critical support for both laboratory and clinical practice, enabling faster and more reliable sample evaluation. The proposed model can accelerate WSI screening by prioritizing histopathology slides with high-risk melanoma features over those with low-risk nevus characteristics. Additionally, it can improve diagnostic accuracy in distinguishing benign from malignant skin lesions, aiding in tumor classification and reducing interobserver variability among pathologists. Overall, this model demonstrates significant potential to advance clinical workflows and diagnostic practices in dermatopathology.

Data availability

The dataset used for the analyses is available from the corresponding author on reasonable request. The code implemented for the automated identification of nuclei and their morphological characterization is publicly available in the Github page of Synthema EU project at https://github.com/synthema-project.

Received: 22 January 2025; Accepted: 16 May 2025
Published online: 01 July 2025

References

1. Sung, H. et al. Global Cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
2. Long, G. V., Swetter, S. M., Menzies, A. M., Gershenwald, J. E. & Scolyer, R. A. Cutaneous melanoma. *Lancet* **402**, 485–502 (2023).
3. Schadendorf, D. et al. *Melanoma the Lancet* **392**, 971–984 (2018).
4. Arnold, M. et al. Trends in incidence and predictions of cutaneous melanoma across Europe up to 2015. *Acad. Dermatol. Venereol.* **28**, 1170–1178 (2014).
5. Nazzaro, G. et al. Dermoscopic predictors of melanoma in small diameter melanocytic lesions (mini-melanoma): a retrospective multicentric study of 269 cases. *Int. J. Dermatology*. **62**, 1040–1049 (2023).
6. Massi, G. & LeBoit, P. E. *Histological Diagnosis of Nevi and Melanoma* (Springer Berlin Heidelberg, 2014). https://doi.org/10.1007/978-3-642-37311-4
7. Bhoyrul, B. et al. Pathological review of primary cutaneous malignant melanoma by a specialist skin cancer multidisciplinary team improves patient care in the UK. *J. Clin. Pathol.* **72**, 482–486 (2019).

8. Hekler, A. et al. Pathologist-level classification of histopathological melanoma images with deep neural networks. *Eur. J. Cancer.* **115**, 79–83 (2019).
9. Shoo, B. A., Sagebiel, R. W. & Kashani-Sabet, M. Discordance in the histopathologic diagnosis of melanoma at a melanoma referral center. *J. Am. Acad. Dermatol.* **62**, 751–756 (2010).
10. Geller, B. M. et al. Pathologists' use of second opinions in interpretation of melanocytic cutaneous lesions: policies, practices, and perceptions. *Dermatol. Surg.* **44**, 177–185 (2018).
11. Gibson, M. et al. Estimating the potential impact of interventions to reduce over-calling and under-calling of melanoma. *Acad. Dermatol. Venereol.* **35**, 1519–1527 (2021).
12. Ronen, S. et al. Discordance in diagnosis of melanocytic lesions and its impact on clinical management. *Arch. Pathol. Lab. Med.* **145**, 1505–1515 (2021).
13. Brochez, L. et al. Inter-observer variation in the histopathological diagnosis of clinically suspicious pigmented skin lesions. *J. Pathol.* **196**, 459–466 (2002).
14. Hekler, A. et al. Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images. *Eur. J. Cancer*. **118**, 91–96 (2019).
15. Massi, D. et al. Impact of second opinion pathology review in the diagnosis and management of atypical melanocytic lesions: A prospective study of the Italian melanoma intergroup (IMI) and EORTC melanoma group. *Eur. J. Cancer*. **189**, 112921 (2023).
16. Kittler, H. Evolution of the clinical, dermoscopic and pathologic diagnosis of melanoma. *Dermatol. Pract. Concept.* **2021163S** https://doi.org/10.5826/dpc.11S1a163S (2021).
17. Elder, D. E., Bastian, B. C., Cree, I. A., Massi, D. & Scolyer, R. A. The 2018 world health organization classification of cutaneous, mucosal, and uveal melanoma: detailed analysis of 9 distinct subtypes defined by their evolutionary pathway. *Arch. Pathol. Lab. Med.* **144**, 500–522 (2020).
18. Lopez-Beltran, A., Canas-Marques, R., Cheng, L. & Montironi, R. Histopathologic challenges: the second OPINION issue. *Eur. J. Surg. Oncol.* **45**, 12–15 (2019).
19. Sauter, D. et al. Validating automatic Concept-Based explanations for AI-Based digital histopathology. *Sensors* **22**, 5346 (2022).
20. Curti, N. et al. Breslow thickness: geometric interpretation, potential pitfalls, and computer automated Estimation. *Pathol. - Res. Pract.* **238**, 154117 (2022).
21. Rashed, H., Flatman, K., Bamford, M., Teo, K. W. & Saldanha, G. Breslow density is a novel prognostic feature in cutaneous malignant melanoma. *Histopathology* **70**, 264–272 (2017).
22. Palve, J. S., Ylitalo, L. K., Luukkaala, T. H., Jernman, J. M. & Korhonen, N. J. A second expert pathology review of cutaneous melanoma in multidisciplinary meetings: impact on treatment decisions. *Surg. Oncol.* **30**, 72–75 (2019).
23. Dika, E. et al. Advantages of manual and automatic computer-aided compared to traditional histopathological diagnosis of melanoma: A pilot study. *Pathol. - Res. Pract.* **237**, 154014 (2022).
24. Xu, H. et al. A whole-slide foundation model for digital pathology from real-world data. *Nature* **630**, 181–188 (2024).
25. Fu, Y. et al. Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. *Nat. Cancer*. **1**, 800–810 (2020).
26. Dixon, A. et al. Online prediction tools for melanoma survival: A comparison. *Acad. Dermatol. Venereol.* **37**, 1999–2003 (2023).
27. Stiff, K. M., Franklin, M. J., Zhou, Y., Madabhushi, A. & Knackstedt, T. J. Artificial intelligence and melanoma: A comprehensive review of clinical, dermoscopic, and histologic applications. *Pigment Cell. Melanoma Res.* **35**, 203–211 (2022).
28. Cazzato, G. et al. Artificial intelligence applied to a first screening of naevoid melanoma: A new use of fast random forest algorithm in dermatopathology. *Curr. Oncol.* **30**, 6066–6078 (2023).
29. Brinker, T. J. et al. Diagnostic performance of artificial intelligence for histologic melanoma recognition compared to 18 international expert pathologists. *J. Am. Acad. Dermatol.* **86**, 640–642 (2022).
30. Cazzato, G. et al. Dermatopathology of Malignant Melanoma in the Era of Artificial Intelligence: A Single Institutional Experience. *Diagnostics* 12, 1972 (2022).
31. Xie, P. et al. Interpretable Classification from Skin Cancer Histology Slides Using Deep Learning: A Retrospective Multicenter Study. Preprint at (2019). https://doi.org/10.48550/ARXIV.1904.06156
32. Sturm, B. et al. Computer-Aided assessment of melanocytic lesions by means of a mitosis algorithm. *Diagnostics* **12**, 436 (2022).
33. Mosquera-Zamudio, A. et al. Deep learning for skin melanocytic tumors in Whole-Slide images: A systematic review. *Cancers* **15**, 42 (2022).
34. Wells, A., Patel, S., Lee, J. B. & Motaparthi, K. Artificial intelligence in dermatopathology: diagnosis, education, and research. *J. Cutan. Pathol.* **48**, 1061–1068 (2021).
35. Schuiveling, M. et al. A novel dataset for nuclei and tissue segmentation in melanoma with baseline nuclei segmentation and tissue segmentation benchmarks. *GigaScience* **14**, giaf011 (2025).
36. Mahbod, A. et al. NuInsSeg: A fully annotated dataset for nuclei instance segmentation in H&E-stained histological images. *Sci. Data*. **11**, 295 (2024).
37. Mahbod, A. et al. CryoNuSeg: A dataset for nuclei instance segmentation of cryosectioned H&E-stained histological images. *Comput. Biol. Med.* **132**, 104349 (2021).
38. Komura, D. et al. Restaining-based annotation for cancer histology segmentation to overcome annotation-related limitations among pathologists. *Patterns* **4**, 100688 (2023).
39. Park, S. et al. Artificial Intelligence–Powered Spatial analysis of Tumor-Infiltrating lymphocytes as complementary biomarker for immune checkpoint Inhibition in Non–Small-Cell lung Cancer. *JCO* **40**, 1916–1928 (2022).
40. Choi, S. et al. Deep learning model improves tumor-infiltrating lymphocyte evaluation and therapeutic response prediction in breast cancer. *Npj Breast Cancer*. **9**, 71 (2023).
41. Helmink, B. A. et al. B cells and tertiary lymphoid structures promote immunotherapy response. *Nature* **577**, 549–555 (2020).
42. Landsberg, J., Tüting, T., Barnhill, R. L. & Lugassy, C. The role of neutrophilic inflammation, angiotropism, and pericytic mimicry in melanoma progression and metastasis. *J. Invest. Dermatology*. **136**, 372–377 (2016).
43. Verma, R. et al. MoNuSAC2020: A Multi-Organ nuclei segmentation and classification challenge. *IEEE Trans. Med. Imaging*. **40**, 3413–3423 (2021).
44. Graham, S. et al. Hover-Net: simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Med. Image. Anal.* **58**, 101563 (2019).
45. Gamper, J. et al. PanNuke Dataset Extension, Insights and Baselines. Preprint at (2020). https://doi.org/10.48550/ARXIV.2003.10778
46. Rahib, L., Wehner, M. R., Matrisian, L. M. & Nead, K. T. Estimated projection of US Cancer incidence and death to 2040. *JAMA Netw. Open*. **4**, e214708 (2021).
47. Postow, M. A., Sidlow, R. & Hellmann, M. D. Immune-Related adverse events associated with immune checkpoint Blockade. *N Engl. J. Med.* **378**, 158–168 (2018).
48. Van Not, O. J. et al. BRAF and NRAS Status and Response to Checkpoint Inhibition in Advanced Melanoma. *JCO Precision Oncology* e2200018 (2022). https://doi.org/10.1200/PO.22.00018
49. Franken, M. G. et al. Trends in survival and costs in metastatic melanoma in the era of novel targeted and immunotherapeutic drugs. *ESMO Open*. **6**, 100320 (2021).
50. Suijkerbuijk, K. P. M., Van Eijs, M. J. M., Van Wijk, F. & Eggermont, A. M. M. Clinical and translational attributes of immune-related adverse events. *Nat. Cancer*. **5**, 557–571 (2024).

51. Van Duin, I. A. J. et al. Baseline tumor-infiltrating lymphocyte patterns and response to immune checkpoint Inhibition in metastatic cutaneous melanoma. *Eur. J. Cancer*. **208**, 114190 (2024).
52. Chatziioannou, E. et al. Deep learning-based scoring of tumour-infiltrating lymphocytes is prognostic in primary melanoma and predictive to PD-1 checkpoint Inhibition in melanoma metastases. *eBioMedicine* **93**, 104644 (2023).
53. Fu, Q. et al. Prognostic value of tumor-infiltrating lymphocytes in melanoma: a systematic review and meta-analysis. *OncoImmunology* **8**, e1593806 (2019).
54. Clark, W. H. et al. Model predicting survival in stage I melanoma based on tumor progression. *JNCI J. Natl. Cancer Inst.* **81**, 1893–1904 (1989).
55. Hendry, S. et al. Assessing Tumor-Infiltrating lymphocytes in solid tumors: A practical review for pathologists and proposal for a standardized method from the international Immuno-Oncology biomarkers working group: part 2: TILs in melanoma, Gastrointestinal tract carcinomas, Non–Small cell lung carcinoma and mesothelioma, endometrial and ovarian carcinomas, squamous cell carcinoma of the head and neck, genitourinary carcinomas, and primary brain tumors. *Adv. Anat. Pathol.* **24**, 311–335 (2017).
56. Busam, K. J. et al. Histologic classification of Tumor-Infiltrating lymphocytes in primary cutaneous malignant melanoma: A study of interobserver agreement. *Am. J. Clin. Pathol.* **115**, 856–860 (2001).
57. Swisher, S. K. et al. Interobserver agreement between pathologists assessing Tumor-Infiltrating lymphocytes (TILs) in breast Cancer using methodology proposed by the international TILs working group. *Ann. Surg. Oncol.* **23**, 2242–2248 (2016).
58. Otsu, N. A threshold selection method from Gray-Level histograms. *IEEE Trans. Syst. Man. Cybern.* **9**, 62–66 (1979).
59. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. (2015). https://doi.org/10.48550/ARXIV.1505.04597
60. Goode, A., Gilbert, B., Harkes, J., Jukic, D. & Satyanarayanan, M. OpenSlide: A vendor-neutral software foundation for digital pathology. *J. Pathol. Inf.* **4**, 27 (2013).
61. Bradski, G. The OpenCV library. *Dr Dobb's J. Softw. Tools* (2000).
62. Haralick, R. M., Shanmugam, K. & Dinstein, I. Textural features for image classification. *IEEE Trans. Syst. Man. Cybern.* **SMC-3**, 610–621 (1973).
63. Rotation and Interpretation of Principal Components. in. *Principal Component Analysis*269–298 (Springer-, 2002). https://doi.org/10.1007/0-387-22440-8_11
64. Hahsler, M., Piekenbrock, M. & Doran, D. Fast Density-Based clustering with *R. J. Stat. Soft.* **91**(1), 1–30 (2019).
65. Diggle, P. J., Tawn, J. A. & Moyeed, R. A. Model-Based geostatistics. *J. Royal Stat. Soc. Ser. C: Appl. Stat.* **47**, 299–350 (1998).
66. Gräler, B., Pebesma, E. & Heuvelink, G. Spatio-Temporal interpolation using Gstat. *R J.* **8**, 204 (2016).
67. Baddeley, A. & Turner, R. Spatstat: an *R* package for analyzing Spatial point patterns. *J. Stat. Soft.* **12**(6), 1–42 (2005).
68. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning* (Springer New York, 2009). https://doi.org/10.1007/978-0-387-84858-7
69. Mardia, K. V., Kent, J. T. & Taylor, C. C. *Multivariate Analysis*vol. 88 (Wiley, 2024).
70. Friedman, J. H. Regularized discriminant analysis. *J. Am. Stat. Assoc.* **84**, 165–175 (1989).
71. Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S* (Springer New York, 2002). https://doi.org/10.1007/978-0-387-21706-2
72. Weihs, C., Ligges, U., Luebke, K. & Raabe, N. KlaR analyzing German business cycles. In *Data Analysis and Decision Support* (eds Baier, D. et al.) 335–343 (Springer, 2005). https://doi.org/10.1007/3-540-28397-8_36.
73. Bienefeld, N. et al. Solving the explainable AI conundrum by bridging clinicians' needs and developers' goals. *Npj Digit. Med.* **6**, 94 (2023).
74. Reddy, S. Explainability and artificial intelligence in medicine. *Lancet Digit. Health*. **4**, e214–e215 (2022).
75. Dika, E. et al. Unraveling the role of MicroRNA/isomiR network in multiple primary melanoma pathogenesis. *Cell. Death Dis.* **12**, 473 (2021).
76. Dika, E. et al. Defining the prognostic role of MicroRNAs in cutaneous melanoma. *J. Invest. Dermatology*. **140**, 2260–2267 (2020).
77. Durante, G. et al. Dysplastic nevi and melanoma: MicroRNAs tell a divergent story. *Pathol. - Res. Pract.* **235**, 153942 (2022).
78. Dika, E. et al. BRAF, KIT, and NRAS mutations of acral melanoma in white patients. *Am. J. Clin. Pathol.* **153**, 664–671 (2020).

Author contributions

G.V. and E.D.: Data Acquisition; G.V., G.Q., and E.D.: Data Validation; A.G. and N.C.: Formal analysis; A.G. and N.C.: Methodology and Software; G.C. and E.D.: Supervision; G.V. and E.D.: Conceptualization; G.C. and E.D.: Project administration; All the authors reviewed the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-02913-z.

Correspondence and requests for materials should be addressed to A.G.