Predictive Nuclear Chromatin Characteristics of Melanoma and Dysplastic Nevi

Matthew G. Hanna[1,2], Chi Liu[3], Gustavo K. Rohde[4,5], Rajendra Singh[2]

[1]Department of Pathology, University of Pittsburgh Medical Center, [3]Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, PA, Departments of [4]Biomedical Engineering and [5]Charles L Brown Electrical and Computer Engineering, University of Virginia, Charlottesville, VA, [2]Department of Pathology and Laboratory Medicine, The Mount Sinai Hospital and Icahn School of Medicine at Mount Sinai, NY, USA

Abstract

Background: The diagnosis of malignant melanoma (MM) is among the diagnostic challenges pathologists encounter on a routine basis. Melanoma may arise in patients with preexisting dysplastic nevi (DN) and it is still the cause of 1.7% of all cancer-related deaths. Melanomas often have overlapping histological features with DN, especially those with severe dysplasia. Nucleotyping for identifying nuclear textural features can analyze nuclear DNA structure and organization. The aim of this study is to differentiate MM and DN using these methodologies. **Methods:** Dermatopathology slides diagnosed as MM and DN were retrieved. The glass slides were scanned using an Aperio ScanScopeXT at ×40 (0.25 μ/pixel). Whole slide images (WSI) were annotated for nuclei selection. Nuclear features to distinguish between MM and DN based on chromatin distributions were extracted from the WSI. The morphological characteristics for each nucleus were quantified with the optimal transport-based linear embedding in the continuous domain. Label predictions for individual cell nucleus are achieved through a modified version of linear discriminant analysis, coupled with the k-nearest neighbor classifier. Label for an unknown patient was set by the voting strategy with its pertaining cell nuclei. **Results:** Nucleotyping of 139 patient cases of melanoma ($n = 67$) and DN ($n = 72$) showed that our method had superior classification accuracy of 81.29%. This is a 6.4% gain in differentiating MM and DN, compared with numerical feature-based method. The distribution differences in nuclei morphology between MM and DN can be visualized with biological interpretation. **Conclusions:** Nucleotyping using quantitative and qualitative analyses may provide enough information for differentiating MM from DN using pixel image data. Our method to segment cell nuclei may offer a practical and inexpensive solution in aiding in the accurate diagnosis of melanoma.

Keywords: Classification, dysplastic nevi, melanoma, nucleotyping, optimal transport

INTRODUCTION

The incidence of melanoma is rising worldwide. The American Cancer Society estimates that 76,100 new cases of melanoma will be diagnosed in the USA and 9710 people are expected to die of metastatic disease.[1] Malignant melanoma (MM) has various risk factors; historically, recurrent severe sun exposure, however environmental and genetic influences also affect the incidence. Dysplastic nevi (DN) may occur sporadically or in a familial setting. In the Caucasian population, sporadic DN are identified in up to 50% of adults.[2,3] DN in the familial setting are estimated around 32,000 in the USA.[4] Patients with DN have an increased risk of developing melanoma.[5] Although patients with DN are at an increased risk for developing melanoma, their role as a precursor lesion is debated. By vast majority, melanomas arise *de novo*. Some studies have shown that up to 25% of melanomas may have an associated dysplastic nevus present, suggesting progression to become melanoma[6,7] although this controversial finding has not been fully substantiated in other studies.

Histologically, DN and melanoma are characterized by their architectural and cytological features. DN are on an average smaller in diameter than melanoma; however, size is not a

Access this article online
Quick Response Code:
Website: www.jpathinformatics.org
DOI: 10.4103/jpi.jpi_84_16

How to cite this article: Hanna MG, Liu C, Rohde GK, Singh R. Predictive nuclear chromatin characteristics of melanoma and dysplastic nevi. J Pathol Inform 2017;8:15.

Available FREE in open access from: http://www.jpathinformatics.org/text.asp?2017/8/1/15/204198

defining factor to distinguish each entity. By definition, DN contain a junctional component, with bridging of adjacent rete ridges. The rete ridges can be uniformly or irregularly elongated. DN have bridging nests, or shouldering, whereas melanoma nests are more coalescent. However, overlapping features can be seen in both entities. Diffuse dermal fibroplasia, associated with a band-like lymphocytic infiltrate, is typically more identified in melanoma, while DN have a more concentric fibroplasia with patchy perivascular lymphocytes. Cytologically, DN should exhibit variable (mild to severe) atypia of the melanocytic cells at the dermal-epidermal junction, while melanoma can show atypia of the dermal component and significant pagetoid spread in epidermis. There is, however, significant morphological overlap between MM and DN, especially in those with severe cytological atypia, with substantial differences in patient prognosis. Pagetoid spread, increased single and confluent melanocytes at the dermal-epidermal junction, and cytological atypia can be seen in both severe DN and melanoma. Mitoses are frequently associated with melanomas, but can be observed in DN. Atypia is defined by nuclear enlargement, irregular nuclear membrane, prominent nucleoli, and coarse chromatin. Many of these histologic features, while on a spectrum, do overlap, which can be difficult to discern each diagnosis.

To complicate matters further, similar genetic mutations have been described in both MM and DN. Two melanoma-susceptibility genes include CDKN2A and BRAF. The presence of DN increases the risk of melanoma in patients with germline mutations in CDKN2A, which is the main genomic locus of melanoma susceptibility.[8] Familial and germline CDKN2A mutations have been reported in DN.[8,9] Somatic mutations of *BRAF* have also been detected in patients with DN and primary melanomas.[10] It has been postulated that BRAF mutations are the initial step in developing melanocytic neoplasia, but alone cannot develop melanoma tumorigenesis.[11]

A lack of consensus between pathologists as well as the National Institutes of Health on the presence of cytological atypia (and architectural disorder) to establish a diagnosis of a dysplastic nevus promotes the need for ancillary tests to aid in diagnosis.[12,13] Dermatologists use diagnostic tools such as the dermascope. In patients with DN, dermascopes have been shown to increase clinical accuracy for diagnosing melanoma. However, the accuracy is highly variable and dependent on expertise and dermoscopy training.[14] Experiments using comparative genomic hybridization (CGH) and fluorescence *in situ* hybridization (FISH) are also being investigated. However, these tests are not without significant cost. CGH is not widely available and requires a 3-week turnaround time. CGH estimates cost between $800 and $2075. FISH studies are more widely available, with a shorter turnaround time, but cost between $1300 and $1800.[15]

Nucleotyping is based principally on chromatin structural changes that are affected by epigenetic changes and DNA binding proteins. DNA organization changes as the cell is in different states. By viewing nuclei in their interphase (and mitotic) state, pixel image data of each nucleus can be transmitted and analyzed. Studies have shown that genomic status (stable vs. unstable) correlates with prognosis, and nucleotyping is being developed to be a potential correlate of analyzing genomic instability.[16,17] Recently, computational methods have attracted much attention toward quantitative analysis of segmented cell or nuclear structures[18-22] and contributed to understanding changes in DNA organization and chromatin structure with regard to cancer development. Nucleotyping including variation in nuclei shape, size, and texture can be analyzed with image analysis pipeline consisting of segmentation, feature extraction, and classification. Employing nucleotyping of textural features uses pixel data derived from the gray-level sum and difference histograms (i.e., co-occurrence matrix) and gray-level run-length matrix. This analysis uses these statistical features to analyze gray levels in the image, their pixel distances, and orientation. Markovian features quantitative gray-level data between adjacent pixels in an image (i.e., run-length, contrast, correlation, difference, entropy, peak, symmetry, etc.). Morphometry features to be analyzed include nuclear size (area, DNA density) and shape (i.e., perimeter, elongation, diameter, cell ferret). Fractal-based methods use a fractal dimension and lacunarity to measure nuclear complexity.[23,24] These methods have been shown to aid in the diagnosis and prognosis of some solid tumors. For instance, nucleotyping was used to develop a biomarker to measure genetic instability in prostate cancer deemed, quantitative nuclear grade (QNG). QNG has been used to predict pathologic stage of disease, recurrence risk after prostatectomy, and also for differentiation of urothelial cell grade in bladder cytology specimens.[25-28] Using multiple morphometric nuclear parameters and retrospective clinical recurrence data, a study showed darker maximum nuclear gray levels and greater variance related to basal cell carcinoma recurrence and worse disease-free survival.[29] Another study has identified nuclear textural features to be an independent prognostic marker of survival in endometrial carcinoma and uterine sarcomas.[30,31] Many studies performing nuclear image analysis to discriminate nuclear textural features have been reportedly successful, however many have not validated on a true independent set of cases. A comprehensive review of nuclear texture analysis showed that <20% of studies at that time evaluated their classifiers on an independent validation set.[32]

In this particular study, we describe a new computational method that differentiates DN and melanoma based on cell nuclear morphology from histopathology images. Melanocyte nuclei in the region of interest (ROI) are manually selected under the supervision of pathologists, which are then segmented automatically. The morphological characteristics for each nucleus are quantified utilizing transport-based morphometry,[21,33] which utilizes mathematics of the optimal

transport (OT) problem in continuous domain,[33] facilitating classification. Figure 1 shows a schematic diagram of the method. In addition to serving as a predictive model, the approach provides visual representation of distribution differences in nuclear morphology between DN and melanoma. If successful, it has the potential to develop new diagnostic, predictive, and prognostic markers for the diagnosis of melanoma, and ultimately advances patient care.

METHODS

Case selection and image acquisition

Under an Institutional Review Board approval, hematoxylin and eosin (H and E) slides from cases of MM and DN were retrieved from the pathology archives. Glass slides of each case were reviewed by two board-certified dermatopathologists. A total of 139 cases were included in our study, including 67 cases diagnosed with MM and 72 cases diagnosed with DN. Slides from each case were scanned with an Aperio ScanScope XT whole slide scanner at ×40 magnification (0.25 μ/pixel) and converted into digital images. The nuclear information embedded in the hematoxylin channel was recovered from RGB images with color deconvolution.[34]

Nuclei segmentation

Melanocyte nuclei were manually annotated and selected for each case. Given nuclei seeds, cell nuclei were segmented automatically using an unsupervised method.[35] In brief, the nucleus edge map in Cartesian-coordinated system is first transformed into the polar space with nucleus seed being the origin. After that, the edge pixels on the nucleus contour are optimally selected by the Dijkstra's algorithm in a constructed shortest path problem. Blurring, overlapping, and incorrect

nuclei segmentations were removed from the dataset by manual inspection. Ultimately, 11,542 melanocyte nuclei were included in the study with 5184 nuclei in the DN group and 6358 nuclei in the MM group. Each nucleus was stored in a 150×150 sub image with black background and bright pixels indicating the amount of locally concentrated chromatin.

Normalization

The intensity summation for each nucleus image was first normalized to be one to minimize the intensity variations in slide preparation, staining procedure, and image acquisition. Next, nuclei position variations such as rotation, translation, and coordinate inversions were eliminated by position normalization. As described in the study by Rohde *et al.*,[36] nuclei were relocated to the image centers to remove translation, and the major axes for nuclei in the dataset were aligned in the same direction to eliminate arbitrary rotation. Figure 2 shows some sample melanocyte nuclei from benign and malignant groups after normalization.

Transport-based morphometry for comparing nuclear chromatin

We utilize the transport-based method described in the studies by Daskalakis *et al.*, Kolouri *et al.*, and Wang *et al.*[22,33,37] for comparing nuclear chromatin. The idea is to use the OT distance to measure the morphological differences between two nuclei, by utilizing all the intensity values. The method consists of two steps: (1) for each nucleus image, compute the transport map with respect to a reference image; (2) utilize the transport map for describing the nucleus morphometry.

As explained in earlier publications,[22,33,37] the method has the advantage that it does not involve information loss, and features can be inverted back to a nucleus image. Thus, classifiers (or any regression method) can be visualized in image space. Here, we provide a brief overview of the method, described in continuous domain:

Given a set of segmented nuclei images, $I_1, I_2, \ldots, I_N: \Omega \rightarrow R$ with $\Omega = [0, 1]^2$, the image intensities are normalized to integrate to one as described above so that the images can be treated as distributions of fixed amount of intensity. One option for the
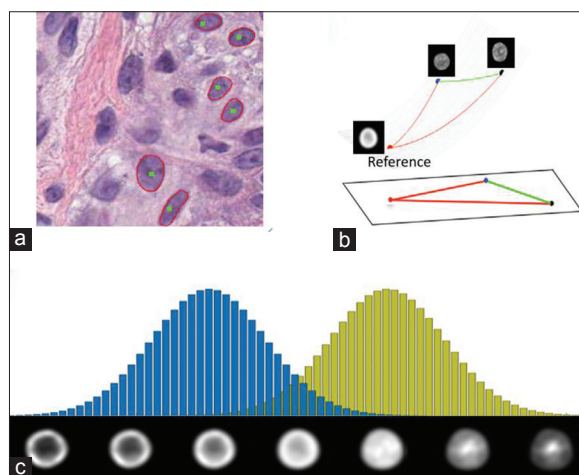

Figure 1: The transport-based nuclei morphometry analysis pipeline. Cell nuclei are first identified by pathologists and segmented from region of interest in the whole slide images (a); morphological characteristics for each individual cell nucleus are represented by the obtained linear embedding utilizing the optimal transport problem (b), which are then used to find the discriminative geodesic for classification and visualization of nuclear distribution differences between malignant melanoma and dysplastic nevi (c)
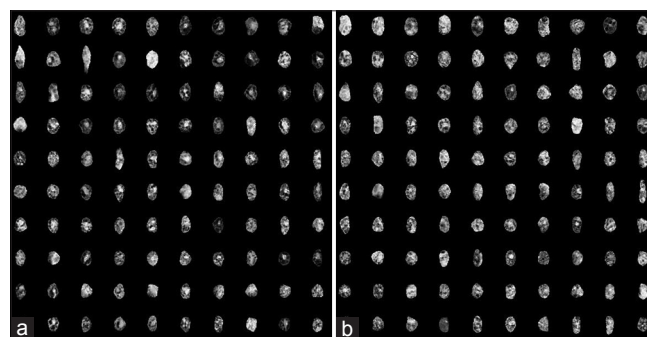

Figure 2: Sample-segmented nuclei from dysplastic nevi (a) and malignant melanoma (b). Note that the intensity of these nuclei images has been inverted (the brighter the pixel, the more chromatin) and the contrast has been stretched for better visualization

choice of reference image is the average of the nuclei image set: $I_0 = \frac{1}{N}\sum_{i=1}^{N} I_i$ and the OT distance between the reference and a given image I_i which is mathematically defined as follows:

$$d_{OT}(I_0, I_i) = \min_{f_i} \int_{\Omega} |f_i(x) - x|^2 I_i(x)\, dx$$

S. t. $\det(Df_i[x])\, I_0(f_i[x]) = I_i(x)$

Where, x is the concatenation of image coordinates including both horizontal and vertical directions. $f_i: \Omega \rightarrow \Omega$ is a mass-preserving transform from I_i to I_0 [Figure 3a]. Df_i is the Jacobian matrix of f_i. $Det(Df_i[x])$ is the determinant of the Jacobian of f_i, as shown in Figure 3b. The transport function f_i is mathematically proved to uniquely exist and can be used to reconstruct I_i from I_0 by $I_i(x) = Det(Df_i[x])\, I_0(f_i[x])$. The morphing process from I_i to I_0 controlled by f_i can be viewed as changing both pixel location and intensity of I_0 at the same time. The points on geodesic f_i connecting I_0 and I_i on the OT manifold can be interpolated by the following formula:

$f\rho(x) = (1 - \rho)x + \rho f_i(x), \rho \in [0, 1]$

The morphing process from I_0 to I_i can be visualized by $\det(DF\rho[x])\, I_0(f\rho[x])$, which is shown in Figure 3c. The OT geodesic f_i can be further decomposed into the identity part x and the OT displacement part $ui(x)$, denoted as $f_i(x) = x + ui(x)$. As described in the study by Kolouri *et al.*,[33] the function $p_i = u_i\sqrt{I_0}$ provides a natural isometric linear embedding for image I_i, which quantifies the morphological characteristics of each nucleus. The linear embedding p_i is the feature representation which can be used for both classification as well as visualization of nuclear distribution differences between benign and malignant groups.

The OT metric not only measures the relative amount of chromatin in certain local regions compared with the reference nucleus image, but also provides a reversible process, where one can reconstruct the nucleus image given the corresponding linear embedding and the reference image [Figure 3].

Classification

To perform nuclei classification, we utilized the standard Fisher linear discriminant analysis (LDA) to construct a one-dimensional subspace which is discriminative for classification. Mathematically, we want to find a linear discriminative vector which can differentiate cell nuclei with their linear embedding. The linear optimal transportation (LOT) embedding for the training dataset can be transformed into such linear subspace, where the class means are well separated, measured relative to the summation of data variances within each class. Given a set of linear embedding, $P = \{p_i\}_{i=1}^{N}$ for nuclei images $I_1, \ldots, I_N$ and their corresponding binary class labels, the projection vector can be computed and denoted as. W_{LDA}^* The projection for nucleus I_i along W_{LDA}^* is computed as $p_i^T W_{LDA}^*$. For any nucleus I_j in the testing set with linear embedding p_j, the projection is similarly obtained as $p_j^T W_{LDA}^*$.
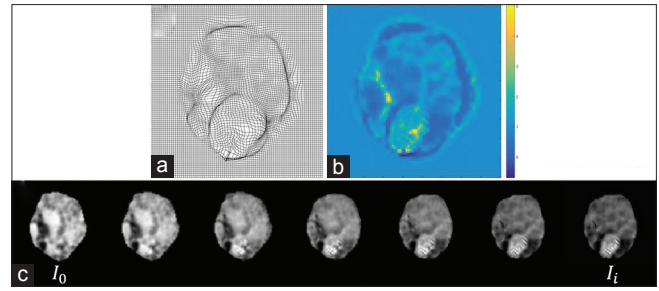

Figure 3: Visualization of the mass-preserving map f_i applied to the image coordinates (a), the determinant of the Jacobian of f_i (b), and the morphing process from I_i to I_0 by changing the parameter $\rho \in [0, 1]$ (c)

The class label for I_j can be predicted through the K-nearest neighbor (KNN) algorithm. The label for a test patient is assigned by the majority voting strategy based on the labels of its pertaining nuclei.[38]

Visualization

The obtained linear embedding for each nucleus can be viewed as an invertible transform. That is, any point p_i in the high-dimensional LOT space can be transformed back to image space through $I_i = \det(Df_i[x])\, I_0(f_i[x])$ with the mass-preserving map f_i. In practice, the number of data points is usually less than the dimension of the LOT space. Thus, principal component analysis (PCA) is first applied to such that the data are represented in a lower dimensional space with the size being the number of the data points. Let the S be the scatter matrix with element S_{kl} being:

$$S_{kl} = \frac{1}{N}\int_{\Omega}(u_k(x) - \bar{u}(x))(u_l(x) - \bar{u}(x) I_0(x)\, dx$$

where, $\bar{u} = \frac{1}{N}\sum_{i=1}^{N} u_i$ is mean displacement filed in the dataset.

The corresponding eigenvalues and eigenvectors can be computed through:

$Se_i = \gamma_i e_i$, $I = 1, \ldots, N$

The PCA direction with regard to can be computed as follows:

$$v_i = \frac{1}{\sqrt{\gamma_i N}}\sum_{k=1}^{N} e_i[k] u_k$$

The mass-preserving map $f^{d,i}$ along the ith prominent direction can be computed through: $f^{d,i} = x + (\bar{u} + d\sigma_i v_i)$, where, σ_i is the standard deviation of data projections on v_i and d is an integer determining the "distance" between the current displacement field and the mean displacement field. Specially, when $d = 0$, $f^{0,i} = x + \bar{u} = \bar{f}$. The reconstructed image corresponding to $f^{d,i}$ can be obtained through: $\det(Df^{d,i}(x))I_0(f^{d,i}(x))$. In the PCA space, each nucleus image I_k is represented by the corresponding linear embedding with reduced dimension:

$$Z_k[i] = \int_{\Omega} v_i(x) u_x(x)\, dx, i = 1, \ldots, N, z^k \in R^{N \times 1}$$

To visualize the discriminative information in the dataset with two classes (benign and malignant in our case), we apply the method described in the study by Wang *et al.*,[39] namely penalized LDA (pLDA), to a set of linear embedding $Z = \{Z_i\}_{i=1}^{N}$ in the PCA-derived space. Given the class label C for each Z_i, the pLDA direction WpLDA onto which the data projections of different classes are separated is obtained by solving the problem below:

$$W_{pLDA} = \arg\max{}_{\|W\|=1} \frac{W^T S_T W}{W^T (S_W + \alpha I) W}$$

Where, $S_T = \sum_i (Z_i - \bar{Z})(Z_i - \bar{Z})^T$ is the scatter matrix and $S_w = \sum_c \sum_{Zi \in c} (z_i - \bar{z}_c)(z_i - \bar{z}_c)^T$ is the within-class scatter matrix with $\bar{Z}c$ being the center of class c. The penalty weight represents the trade-off between the standard LDA and PCA directions.[38] In the experiment, was selected by measuring the stability of the pLDA subspace.[22] The discriminative direction in LOT embedding space is given by: $v_{pVDA} = \sum_{K=1}^{N} W*[k] u_k$. By changing the integer d, the discriminative information along V_{pLDA} around the mean displacement field can be visualized by:
$$\det\left(Df_{pLDA}^d (x)\right) I_0 \left(f_{pLDA}^d (x)\right), \text{where } f_{pLDA}^d = x + (\bar{u} + d\sigma v_{pVDA})$$
and σ is the standard deviation after data projection along V_{pLDA}.

RESULTS

To evaluate the performance of the proposed method, we predicted the class labels (DN or MM) for test patients after the computation of linear embedding, representing the morphology of segmented nuclei and then compared the predicted labels with actual diagnoses. Our experiment utilized the standard leave-one-patient-out validation scheme, whereby cell nuclei pertaining to one patient were removed from the dataset and kept as testing data. In the training stage, the training data were further split into training and validation sets to search for the optimal k in the KNN algorithm. The one-dimensional pLDA subspace was constructed using the LOT embedding and class labels in the training data. In our case, the discriminative direction was computed and applied to obtain the projections of the testing data samples. The binary labels for individual testing nucleus were given by the KNN algorithm, and the label for any test patient was assigned through the majority voting strategy (taking the most common label assignment among its pertaining nuclei).

For comparison purpose, we also tested the performance of the numerical feature-based method described in a study by Basu *et al.*[21] The morphology of each nucleus was represented using a 256-dimensional numerical feature vector, including six morphological features (area, convexity, circularity, perimeter, eccentricity, and equivalent diameter), 220 texture features (intensity-based features, Haralick features, and Gabor features), and thirty wavelet features. Similar to the classification procedure for OT-based method, the LDA

technique was first applied to the derived feature matrix to find the discriminative direction, which was then used in label prediction for test patients together with the KNN algorithm and majority voting strategy. Table 1 shows a summary of classification results by our proposed method and the numerical feature-based method. As we can see from the confusion matrices table, our LOT-based method is clearly superior compared with numerical feature-based method with 6.4% gain [Table 1].

Our method enables direct visualization of the discriminant information embedded in cell nuclei from benign and malignant groups. Visual comparisons of nuclei morphology for DN versus MM are shown in Figure 4. The nuclei morphological information in the held out data (30% of data) was visually plotted along the first pLDA direction, computed based on linear embedding in the PCA space and corresponding labels in the training set. In Figure 4, the height of each histogram bar indicates the proportion of nuclei most similar to the representative image beneath it. The row of representative images is the visual representation along the discriminant geodesic path by pLDA. These representatives are generated between every unit of standard deviation from the mean image along the derived direction V_{pLDA}. The P value for differentiating data points between benign and malignant groups after projection is zero within numerical precision, indicating that DN and MM are significantly different in nuclei morphological distribution [Figure 4].

As far as biological interpolation shown in Figure 4, the most important information in discriminating DN and MM is the chromatin distribution patterns inside the nuclei as well as the nuclei size. In comparison with MM, it is uncommon for DN

Table 1: Confusion matrices for linear optimal transportation-based method (in bold) and numerical feature-based method

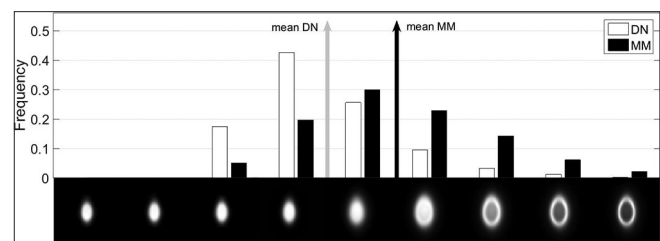|	Predicted Diagnosis	
	DN	**MM**
Actual diagnosis		
DN	61/57	11/15
MM	15/20	52/47
Accuracy	81.29%/74.82%	

Figure 4: Histograms of nuclei morphological distributions for dysplastic nevi (white bars) and malignant melanoma (dark bars). Note that the intensity of the representative images beneath the histogram bars has been inverted with brighter pixels indicating more locally accumulated chromatin

nuclei to have a chromatin distribution centered at the nuclei periphery; on the contrary, nuclei of DN tend to have more uniform chromatin distributions. To show this, we measured the ratio between mean intensity of pixels around the nuclei membrane and mean intensity of the remaining central area inside the nuclei and plotted ratio distributions for both DN and MM groups in Figure 5. As one can see from the ratio distributions, MM nuclei tend to show higher proportions in high membrane/center ratios compared to the nuclei in DN group, which corresponds with the discriminating nuclei patterns between DN and MM in Figure 4. In a study by Massi and LeBoit[40] atypical melanocyte nuclei in MM are larger and the chromatin is coarser, with an irregular nuclear membrane, while stereotypical benign melanocytes have evenly distributed chromatin and rounder nuclear membrane. From the perspective of cell biology, the DNA exists in the nucleus in two forms: heterochromatin and euchromatin, reflecting the level of activity of cell. Heterochromatin is darkly stained (bright if the intensity is inverted) by H and E staining, usually scattered throughout the nucleus or adjacent to the nuclear envelope, and is typical of quiescent nuclei. Euchromatin, on the contrary, is not readily stainable and is prevalent in cells that are active in gene transcription.[41] Therefore, our finding agrees with the fact that benign cell nuclei appear more heterochromatic and malignant cell nuclei appear more euchromatic, and thus the approximately "ring shape" nuclear appearance in MM relative to the nuclei in DN [Figure 5].

In addition, as observed from Figure 4, a greater number of MM nuclei are relatively bigger than the nuclei in DN group. We also plotted the histogram of nuclei size distribution for both DN and MM groups, as shown in Figure 6. The average nuclei size is 647.04 (in pixels) for MM group, compared with the average nuclei size of 543.07 in DN group, demonstrating that our proposed method can capture and visualize the distribution differences in nuclei size between the two classes [Figure 6].

DISCUSSION

Melanocytic lesions are among the most challenging areas in surgical pathology and are a significant proportion of dermatopathology daily case volume. The lack of consensus between pathologists, significant morphological overlap as
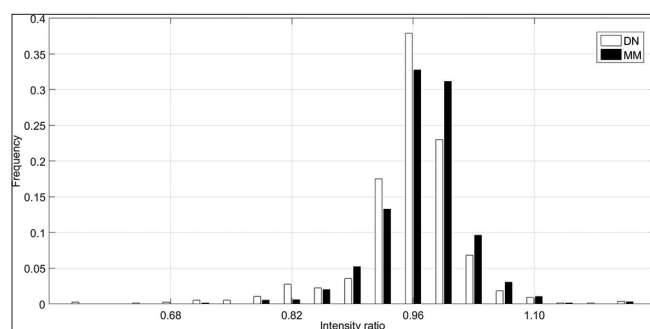
well as similar gene mutations between MM and DN make clinical diagnosis complicated.

We have described a novel image analysis method to quantify the characteristics of nuclei morphology to differentiate benign (DN) and MM melanocytic lesions. The cell nuclei were segmented from manually located ROI in H and E stained WSI. The LOT-based method then describes the characteristics of morphology of each nucleus by the linear embedding for classification/quantitative analysis. The proposed framework is based on the linear OT in continuous domain, together with the simple majority voting strategy with individual nucleus label predicted by KNN algorithm after discriminative projections. With nuclei morphological information alone, the method showed an accuracy of 81.3%, an improved classification with 6.4% gain in differentiating patients from benign or malignant in double cross-validations, compared with the commonly used numerical features. In addition, the distribution differences in nuclei morphology can be directly visualized along the discriminative geodesic generated by the penalized version of LDA (pLDA). In the analysis of large case samples (139 patients in our study), the proportions of representative nuclei along the pLDA direction for DN and MM are statistically different with meaningful biological interpretations, which agrees well with descriptions in the existing literatures and actual morphological measurements.

We also note that the potential reasons behind the improved classification accuracy of LOT in our study can be explained in two folds: (1) LOT performed in continuous domain is a reversible transformation. In other words, theoretically, one can reconstruct the original nucleus image without information loss given its corresponding linear embedding and the reference image. Such property enables the preservation of discriminative information when quantifying nuclei morphological characteristics for classification. On the contrary, numerical feature-based method usually prevents reconstruction of the input images from feature space, and the performance boosting requires more manually designed discriminative features involved to represent the nucleotyping. (2) The transformation from image domain to the LOT space is a nonlinear operation. As explained in the study by Rohde
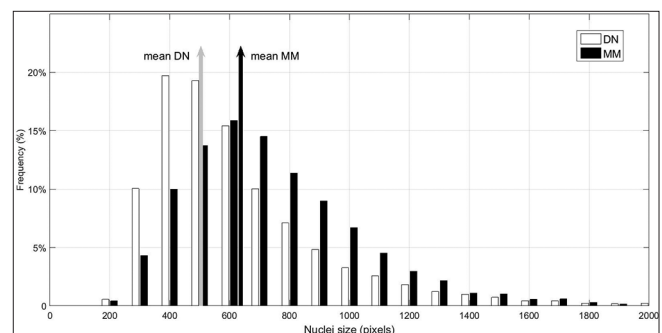

Figure 5: Histograms of membrane/center ratio distributions for dysplastic nevi (white bars) and malignant melanoma (black bars)

Figure 6: Histograms of nuclei size distributions for dysplastic nevi (white bars) and malignant melanoma (black bars). The mean sizes for nuclei from dysplastic nevi and malignant melanoma are 543.07 pixels and 647.04 pixels, respectively

et al.,[36] the nonlinearity has the potential to improve linear separability, and thus improves the classification accuracy in our study. Feature-based method, however, often contains a mixture of linear and nonlinear operations depending on the type of measurements we are interested in.

In the future, our work can be extended in the following aspects: first, fully automated and reliable nuclei detection and segmentation algorithms are required when a larger number of cases are involved in the study. WSI often constitutes of a mosaic of different kinds of cells (e.g., melanocytes and keratinocytes in our case) in various tissue types, which poses a huge challenge when analyzing a certain kind of cells in the nuclei-based cancer detection pipeline. Even if the nuclei of interest are located by pathologists, to the best of our knowledge, existing segmentation methods produce segmentation failures with different degrees, which may affect the classification performance when manually screening segmentations is impossible for large datasets. Second, alternative nonnaive Bayesian models need to be developed and applied to predict patient labels. In our experiment, patient-level labels are assigned through the majority voting strategy after predicting cell nuclei labels individually and independently. Though implicitly, such procedure actually follows the naive Bayesian model, where there is no assumed interdependency among cell nuclei. As shown in a study by Huang *et al.*,[42] neglecting interdependencies between nearby nuclei can lead to suboptimal classification results.

Finally, our study shows that nucleotyping alone can obtain an accuracy of 81.3%. This suggests a practical, inexpensive solution, with similar accuracy like the current clinical dermascopes and molecular diagnostics. With further development and refined analysis, nucleotyping may be of prime use as a screening tool. Large data sets of digital images can be automatically screened and flagged for review as necessary. Ongoing research of new diagnostic characteristics can be used in combination with other modalities for increased accuracy in the diagnosis and prognosis of patients.

CONCLUSIONS

Our study visually identifies the distribution differences between MM and DN based on novel representations of nuclear morphology. Using WSI and the proposed LOT-based nuclei analysis framework, MM and DN are distinguishable with a better accuracy compared with numerical feature-based method. Nuclear image analysis in conjunction with next-generation sequencing or other feature-rich detection systems may increase diagnostic accuracy and has prognostic implications. The clinical and functional significance of nuclear chromatin morphologies in cancer has yet to be fully realized. Larger studies with combinations of other features and survival data are needed to draw further diagnostic and prognostic conclusions.

Financial support and sponsorship

Conflicts of interest
There are no conflicts of interest.

REFERENCES

1. Schadendorf D, Fisher DE, Garbe C, Gershenwald JE, Grob JJ, Halpern A, *et al.* Melanoma. Nat Rev Dis Primers 2015;1:15003.
2. Cooke KR, Spears GF, Elder DE, Greene MH. Dysplastic naevi in a population-based survey. Cancer 1989;63:1240-4.
3. Crutcher WA, Sagebiel RW. Prevalence of dysplastic naevi in a community practice. Lancet 1984;1:729.
4. Kraemer KH, Greene MH, Tarone R, Elder DE, Clark WH Jr., Guerry D 4th. Dysplastic naevi and cutaneous melanoma risk. Lancet 1983;2:1076-7.
5. Naeyaert JM, Brochez L. Clinical practice. Dysplastic nevi. N Engl J Med 2003;349:2233-40.
6. Marks R, Dorevitch AP, Mason G. Do all melanomas come from "moles"? A study of the histological association between melanocytic naevi and melanoma. Australas J Dermatol 1990;31:77-80.
7. Elder DE. Dysplastic naevi: An update. Histopathology 2010;56:112-20.
8. Goldstein AM, Martinez M, Tucker MA, Demenais F. Gene-covariate interaction between dysplastic nevi and the CDKN2A gene in American melanoma-prone families. Cancer Epidemiol Biomarkers Prev 2000;9:889-94.
9. Bishop JA, Wachsmuth RC, Harland M, Bataille V, Pinney E, MacK P, *et al.* Genotype/phenotype and penetrance studies in melanoma families with germline CDKN2A mutations. J Invest Dermatol 2000;114:28-33.
10. Friedman RJ, Farber MJ, Warycha MA, Papathasis N, Miller MK, Heilman ER. The "dysplastic" nevus. Clin Dermatol 2009;27:103-15.
11. Pollock PM, Harper UL, Hansen KS, Yudt LM, Stark M, Robbins CM, *et al.* High frequency of BRAF mutations in nevi. Nat Genet 2003;33:19-20.
12. Diagnosis and treatment of early melanoma. NIH Consensus Development Conference. January, 27-29. Consens Statement 1992;10:1-25.
13. Clark WH Jr., Evans HL, Everett MA, Farmer ER, Graham JH, Mihm MC Jr., *et al.* Early melanoma. Histologic terms. Am J Dermatopathol 1991;13:579-82.
14. Kittler H, Pehamberger H, Wolff K, Binder M. Diagnostic accuracy of dermoscopy. Lancet Oncol 2002;3:159-65.
15. March J, Hand M, Truong A, Grossman D. Practical application of new technologies for melanoma diagnosis: Part II. Molecular approaches. J Am Acad Dermatol 2015;72:943-58.
16. Kristensen GB, Kildal W, Abeler VM, Kaern J, Vergote I, Tropé CG, *et al.* Large-scale genomic instability predicts long-term outcome for women with invasive stage I ovarian cancer. Ann Oncol 2003;14:1494-500.
17. Hveem TS, Merok MA, Pretorius ME, Novelli M, Bævre MS, Sjo OH, *et al.* Prognostic impact of genomic instability in colorectal cancer. Br J Cancer 2014;110:2159-64.
18. Chen C, Wang W, Ozolek JA, Rohde GK. A flexible and robust approach for segmenting cell nuclei from 2D microscopy images using supervised learning and template matching. Cytometry A 2013;83:495-507.
19. Gurcan MN, Boucheron LE, Can A, Madabhushi A, Rajpoot NM, Yener B. Histopathological image analysis: A review. IEEE Rev Biomed Eng 2009;2:147-71.
20. Wang W, Ozolek JA, Rohde GK. Detection and classification of thyroid follicular lesions based on nuclear structure from histopathology images. Cytometry A 2010;77:485-94.
21. Basu S, Kolouri S, Rohde GK. Detecting and visualizing cell phenotype differences from microscopy images using transport-based morphometry. Proc Natl Acad Sci U S A 2014;111:3448-53.
22. Daskalakis A, Kostopoulos S, Spyridonos P, Glotsos D, Ravazoula P, Kardari M, *et al.* Design of a multi-classifier system for discriminating benign from malignant thyroid nodules using routinely H and E-stained cytological images. Comput Biol Med 2008;38:196-203.

23. Nielsen B, Albregtsen F, Danielsen HE. Fractal signature vectors and lacunarity class distance matrices to extract new adaptive texture features from cell nuclei. In: Fractals in Biology and Medicine. Vol. 3. Basel: Birkhäuser-Verlag; 2002. p. 55-65.

24. Nielsen B, Albregtsen F, Kildal W, Danielsen HE. Prognostic classification of early ovarian cancer based on very low dimensionality adaptive texture feature vectors from cell nuclei from monolayers and histological sections. Anal Cell Pathol 2001;23:75-88.

25. Veltri RW, Partin AW, Miller MC. Quantitative nuclear grade (QNG): A new image analysis-based biomarker of clinically relevant nuclear structure alterations. J Cell Biochem Suppl 2000;Suppl 35:151-7.

26. Veltri RW, Miller MC, Mangold LA, O'Dowd GJ, Epstein JI, Partin AW. Prediction of pathological stage in patients with clinical stage T1c prostate cancer: the new challenge. J Urol 2002;168:100-4.

27. Veltri RW, Miller MC, Isharwal S, Marlow C, Makarov DV, Partin AW. Prediction of prostate-specific antigen recurrence in men with long-term follow-up postprostatectomy using quantitative nuclear morphometry. Cancer Epidemiol Biomarkers Prev 2008;17:102-10.

28. Wojcik EM, Miller MC, O'Dowd GJ, Veltri RW. Value of computer-assisted quantitative nuclear grading in differentiation of normal urothelial cells from low and high grade transitional cell carcinoma. Anal Quant Cytol Histol 1998;20:69-76.

29. Cheretis C, Angelidou E, Dietrich F, Politi E, Kiaris H, Koutselini H. Prognostic value of computer-assisted morphological and morphometrical analysis for detecting the recurrence tendency of basal cell carcinoma. Med Sci Monit 2008;14:MT13-19.

30. Nielsen B, Hveem TS, Kildal W, Abeler VM, Kristensen GB, Albregtsen F, *et al.* Entropy-based adaptive nuclear texture features are independent prognostic markers in a total population of uterine sarcomas. Cytometry A 2015;87:315-25.

31. Geisler JP, Wiemann MC, Zhou Z, Miller GA, Geisler HE. Markov texture parameters as prognostic indicators in endometrial cancer.

Gynecol Oncol 1996;62:174-80.

32. Nielsen B, Albregtsen F, Danielsen HE. Statistical nuclear texture analysis in cancer research: A review of methods and applications. Crit Rev Oncog 2008;14:89-164.

33. Kolouri S, Tosun AB, Ozolek JA, Rohde GK. A continuous linear optimal transport approach for pattern analysis in image datasets. Pattern Recognit 2016;51:453-62.

34. Ruifrok AC, Johnston DA. Quantification of histochemical staining by color deconvolution. Anal Quant Cytol Histol 2001;23:291-9.

35. Liu C, Shang F, Ozolek JA, Rohde GK. Detecting and segmenting cell nuclei in 2D microscopy images. J Pathol Inform 2016;7:42.

36. Rohde GK, Ribeiro AJ, Dahl KN, Murphy RF. Deformation-based nuclear morphometry: capturing nuclear shape variation in HeLa cells. Cytometry A 2008;73:341-50.

37. Wang W, Slepcev D, Basu S, Ozolek JA, Rohde GK. A linear optimal transportation framework for quantifying and visualizing variations in sets of images. Int J Comput Vis 2013;101:254-69.

38. Tosun AB, Yergiyev O, Kolouri S, Silverman JF, Rohde GK. Detection of malignant mesothelioma using nuclear structure of mesothelial cells in effusion cytology specimens. Cytometry A 2015;87:326-33.

39. Wang W, Mo Y, Ozolek JA, Rohde GK. Penalized fisher discriminant analysis and its application to image-based morphometry. Pattern Recognit Lett 2011;32:2128-35.

40. Massi G, LeBoit PH. Histological Diagnosis of Nevi and Melanoma. Berlin: Springer Science and Business Media; 2013.

41. Frenster JH. Ultrastructure and function of heterochromatin and euchromatin. In: Busch H, editor. The Cell Nucleus. Vol. 1. New York: Academic Press; 1974. p. 565-81.

42. Huang H, Tosun AB, Guo J, Chen C, Wang W, Ozolek JA, *et al.* Cancer diagnosis by nuclear morphometry using spatial information. Pattern Recognit Lett 2014;42:115-21.