

Article

Staging Melanocytic Skin Neoplasms Using High-Level Pixel-Based Features

Mai Ramadan Ibraheem ¹, Shaker El-Sappagh ^{2,3}, Tamer Abuhmed ^{4,*} 
and Mohammed Elmogy ⁵ 

¹ Information Technology Department, Faculty of Computers and Information, Kafrelsheikh University, Kafrelsheikh 33516, Egypt; Mai_Amer@fci.kfs.edu.eg

² Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS), Universidade de Santiago de Compostela, 15705 Santiago de Compostela, Spain; shaker.elsappagh@usc.es

³ Information Systems Department, Faculty of Computers and Artificial Intelligence, Benha University, Benha 13512, Egypt

⁴ Department of Computer Science and Engineering, College of Computing, Sungkyunkwan University, Seoul 06351, Korea

⁵ Information Technology Department, Faculty of Computers and Information, Mansoura University, Mansoura 35516, Egypt; melmogy@mans.edu.eg

* Correspondence: tamer@skku.edu

Received: 6 August 2020; Accepted: 2 September 2020; Published: 4 September 2020



Abstract: The formation of malignant neoplasm can be seen as deterioration of a pre-malignant skin neoplasm in its functionality and structure. Distinguishing melanocytic skin neoplasms is a challenging task due to their high visual similarity with different types of lesions and the intra-structural variants of melanocytic neoplasms. Besides, there is a high visual likeliness level between different lesion types with inhomogeneous features and fuzzy boundaries. The abnormal growth of melanocytic neoplasms takes various forms from uniform typical pigment network to irregular atypical shape, which can be described by border irregularity of melanocyte lesion image. This work proposes analytical reasoning for the human-observable phenomenon as a high-level feature to determine the neoplasm growth phase using a novel pixel-based feature space. The pixel-based feature space, which is comprised of high-level features and other color and texture features, are fed into the classifier to classify different melanocyte neoplasm phases. The proposed system was evaluated on the PH2 dermoscopic images benchmark dataset. It achieved an average accuracy of 95.1% using a support vector machine (SVM) classifier with the radial basis function (RBF) kernel. Furthermore, it reached an average Disc similarity coefficient (DSC) of 95.1%, an area under the curve (AUC) of 96.9%, and a sensitivity of 99%. The results of the proposed system outperform the results of other state-of-the-art multiclass techniques.

Keywords: pigmented skin lesions; pixel-based features; high-level features; pigment network; melanocyte neoplasm phases; globules and streaks

1. Introduction

Pigmented skin lesions represent about 20% of all skin cancer cases [1]. Pigmented skin lesions are generally divided into melanocytic and non-melanocytic lesions [2]. Melanocytic lesions refer to lesions that have different colors, most often due to melanin, such as Melanocyte nevi, Solar Lentigo, Dermatofibromas (DFs), and Vascular lesions (VASC) [3]. Non-melanocytic lesions lack melanin pigment. Their color affects other factors, i.e., hemoglobin or keratin, such as keratinocytic, VASC, and reactive lesions [4]. Exposure to ultraviolet radiation from the sun increases the risk of these lesions to be malignant pigmented lesions or skin cancer.

Skin cancer can be commonly classified as melanoma and non-melanoma pigmented lesions [4]. Melanoma is related to melanocytes, changing the color of malignant cells. The most popular forms of non-melanoma skin cancer can be squamous cell carcinoma (SCC) and basal cell carcinomas (BCC). Non-melanoma is commonly popular than melanoma, but melanoma accounts for most mortality of pigmented skin [5].

Determination of unknown lesion of the skin is significant to specify the suitable treatment. The curability of skin cancer has higher rates if it is detected early enough and cured surgically [6]. Due to the complexity of skin cancer treatment at later stages, the investigation of an efficient non-invasive automated system can help in guiding diagnosis. As skin cancer diagnosis faces many problems, including human faults and higher cost, researchers attempt to automate the diagnosis process to verify it is benign or dangerous with high accuracy [3]. Therefore, high-performance computer-assisted diagnostic (CAD) systems can help guide the diagnosis.

Melanocytic pigmented lesions of the skin, i.e., Melanocytic nevi (Melcyt NV), can progress to pre-malignant dysplastic nevi (Dysp NV) or malignant melanoma (Mel) [7], as shown in Figure 1. Thus, patients can be survived in early detected cases [8]. Early detection is a promising strategy to reduce the skin cancer mortality rate [7]. The ability to rapidly spread to other parts of the body is a unique feature that makes Melanoma one of the deadly diseases of the skin [8].

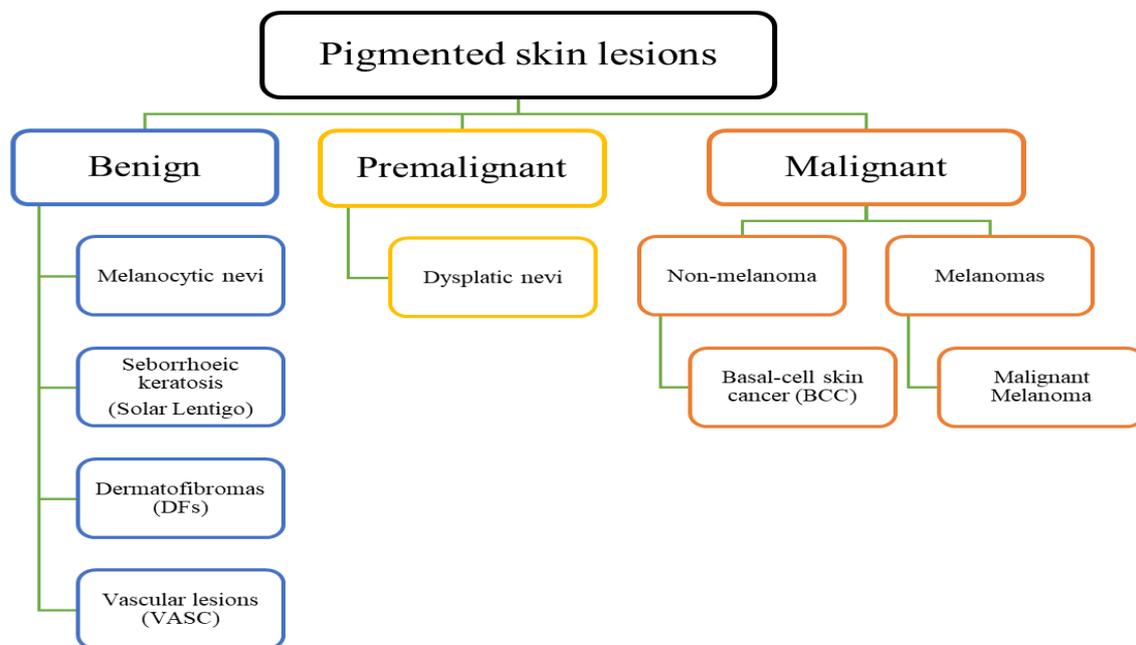


Figure 1. The pigmented skin lesions at different stages.

The pigmented network can be typical or atypical [7]. The typical pigmented network is a regular pigment network, which has uniform brown lines and rete ridges. These ridges are relatively similar in width and equidistant as in Melcyt NV, and non-melanocytic lesions, such as lentigo and dermatofibromas [9]. The pigment network comprises intersecting brown lines in the form of a grid pattern, as shown in Figure 2 [10]. These brown lines refer to disorders of pigmentation either within the keratinocytes or the melanocytes.



Figure 2. The typical pigmented network.

Atypical pigment network has irregular lines varying in size, color, thickness, or melanin distribution, often found in Dysp NV [10], as shown in Figure 3. Globules have a round to oval shape located along the borders of a melanocytic lesion and commonly found in a growing nevus [11], as shown in Figure 4.



Figure 3. The atypical pigmented network.



Figure 4. The globule dermoscopic structures.

Streaks are linear pigmented within the borders of a melanocytic lesion and include radial flowing (lineal streaks) and round projections (pseudopods). The symmetrical distribution of streaks along the margin of the melanocytic lesion favors the prognosis of a nevus. Still, asymmetrical distribution can be expected to lead to the spread of melanoma [12], as shown in Figure 5.



Figure 5. Streaks and globules dermoscopic structures.

However, the significant similarity between peripheral globules and pseudopods is that the bordering globules are small clear spaces separating the globule from the primary tumor mass. They correspond to melanocytic cells and are usually associated with growing nevi. The radial linear extends from the periphery of the lesion in a process known as radial streaming. In histopathology, radial streaming also corresponds to pigmented melanocytes [12]. In contrast, pseudopods are directly connected to the primary tumor mass through a stem, which favors the diagnosis of melanoma. Thus, discrimination between radial lines and pseudopods can guide the diagnosis of malignant melanoma from melanocytic nevi [13].

Thus, circumferentially distributed radial lines characterize the presence of Nevi, but segmental pseudopods most likely represent the occurrence of melanoma. According to the clinical description, malignant lesions are characterized by distinct high borders compared to benign cases [5]. The malignant lesion has a different structure than the benign lesion. Thus, the irregularity of lesion borders can provide intuitive rationale feature space for the judgment of melanoma.

Earlier dermatoscopic techniques, i.e., the asymmetry, border, color, and diameter (ABCD) rule, seven-point checklist, Menzies, and CASH algorithm [14], focused on determining lesion type by determining the lesion's area and recognizing its various features. Recent skin imaging techniques enable the visualization of skin structures, which positively affects the diagnostic capability and enhance results [11]. The distributions of color and texture features also provide good discrimination of pigmented skin lesions from unaffected skin regions in the image.

Color features are essential in the assessment of dermoscopic lesions. Most skin colors are caused by an increase in a given chromophore, i.e., melanin is the most critical chromophore in pigmented lesions. These colors may be brown, black, gray, or blue for pigment; yellow for lipids or keratin; white for collagen; or red for blood [4].

Texture analysis attempts to identify, measure, and detect the differences between different regions. The texture can be measured for a region but cannot be detected for a single point. The textural features analysis methods can be divided into statistical and structural methods. Statistical methods look for pixel value relationships and statistical moments, while structural methods are concerned more with regions such as shapes and edges [15].

Statistical methods describe point pixel properties based on gray level statistical moments, using a co-occurrence matrix or extended species distribution models (SDM) [16]. Statistical methods use local grey level statistics to define texture, which is constant or varies slowly over a textured region [17].

Feature descriptors are generally used for capturing unique metrics from whole images or image regions, including textural, statistical, model-based, and basis space methods [14]. Feature descriptors vary in capturing technique and their attributes. Feature detectors determine interesting features in the image, such as interest point, keypoint, or landmark. The key point detector determines the vector orientation of the neighborhood feature descriptor, which provides some amount of invariance and robustness [18].

The shape descriptor concerns the measuring of several pixel regions among the shapes to be used for descriptor computations [19]. The morphological boundary shape descriptor is a method for defining polygon and boundary shape. Morphological shapes are generally described as blobs.

Thresholding is often the first step in defining object boundaries [18]. Morphological reshape operators clean up the shape boundary by growing or shrinking, using erode and dilate techniques [5].

High-level features are features that have been designed using human-observable formulation models, contrary to low-level features that were designed without describing human-observable formulation models. Integration between high-level and low-level features increases the significance of the feature space, which allows the system to provide a more reasonable justification for the classification decision [20].

The feature space that contains high-level features can describe the other characteristics of melanoma, i.e., the border shape and structure of patterns. Implementing a non-invasive, automated pigmented skin lesion system able to identify the type of the lesion could save lives and reduce unnecessary biopsies, in addition to cost reduction [21].

For this goal, this work develops a feature extraction technique-based greyscale, texture morphology, statistical area filters, and basis space filtering for detecting pigment network. The association between several descriptors for the same object increases discrimination rates.

The objectives of this paper are to develop a staging framework that could regulate the progression of skin lesions and evaluate the use of concerning features identifying pigmented lesions accurately, ranging from a Melcyt NV and Dysp NV to Mel. The feature space, which comprises of low-level features for local region pixel details and high-level features for regional shape metrics, were extracted in a pixel passed manner. The segmentation of melanocytic neoplasms is also conducted in a pixel passed manner.

The extracted feature space is used as inputs to different classifiers to distinguish different phases of melanocyte neoplasm. The proposed system was evaluated on the PH2 dermoscopic images benchmark dataset. The high-level pixel-based features suggested are considered reliable biomarkers for melanoma diagnosis and achieved an average accuracy of 95.1% using the support vector machines (SVM) classifier with the radial basis function (RBF) kernel.

The rest of this paper is organized as follows: Section 2 presents various dermatoscopic techniques for several studies. Section 3 introduces the proposed framework for staging melanocytic neoplasms using high-level pixel-based features. Section 4 describes the dataset used and experimental results. Finally, Section 5 concludes the work presented.

2. Related Work

Most of the recently evolved dermoscopic algorithms were developed to facilitate the ability to distinguish different types of melanocytic neoplasms, using the ABCD rule and low-level features [22–25]. However, recent studies suggest incorporation between high- and low-level descriptors is a sign that refers to lesion borders. For example, Gutman et al. [26] demonstrated automatic detection for globule and streak dermoscopic features. They performed localization and classification based on superpixels and dermoscopic features in order to judge the presence and absence of the globules and streaks. They evaluated the results using 807 images of training data and the 335 testing datasets from the international skin Imaging collaboration (ISIC) 2016 dataset, which has a classification accuracy of 91%. ISIC 2016 lacks an intuitive mapping label for globule and streak dermoscopic features for further analysis. Do et al. [27] adopted a melanoma detection system to localize lesions of the skin using a set of image features and a hierarchical approach for segmentation. They computed the border irregularity of the shape features, i.e., convexity, compactness, and distance variance between border points and lesion centroid. They evaluated the results on images obtained from the Singapore National Skin Center, comprised of 117 benign nevi and 67 malignant melanomas, and achieved 89.09% sensitivity and 90% specificity. Pixel-based feature extraction techniques can lead to a better distribution of lesion features to avoid misleading results. Lee et al. [28] designed a multiclass skin disease classification system. Their model comprises of segmentation based (DenseNet and U-net) pre-processing steps fed into a successive fine-tune of classification models. They performed the classification of seven skin diseases to predict the disease class. They achieved results using the HAM10000 Dataset, obtaining

classification accuracies of 0.899 and 0.785. The proposed model can be planned to add more flexibility to the system. Abbadi and Faisal [22] presented an automated skin image diagnosis based on ABCD rules and a new asymmetry determination method. They computed asymmetry by dividing the lesion into horizontal and vertical parts, then counted the number of mismatched pixels between the two parts using the union and intersection of the two parts. The proposed method was tested on 220 different images of 120 images from the PH2 database and 100 images from the websites, in which 113 images are of cancer, and 107 images are of non-cancer, achieving an accuracy of 95.45%. They tested accuracy only on a sample of malignant and benign lesion images, in which the PH2 dataset included multiple stages of skin lesion. Nammalwar et al. [29] integrated both texture and color features for segmenting lesions of the skin. They used ABCD's clinical features of pigmented lesions as a measurement to detect and localize lesions in skin images, i.e., characteristics of the color, maximum of diameter, and irregularity of the boundary. They first extracted texture and color information to be used in segmenting lesion boundaries. Modified Kolmogorov-Smirnov (MKS) was used to discriminate the texture distribution to feed into a boundary refinement algorithm in order to obtain the final segmented image. They evaluated the proposed model on 18 skin cancer images obtained from the dermatology gallery, based on the comparison with the Live Wire segmentation results. The author's main concern was to extract significant features for the segmentation method. They did not provide any quantitative comparisons that assure the accuracy of their proposed method. Codella et al. [30] proposed a system for the classification of melanoma using skin dermoscopic images. They combined the more recent machine learning techniques, deep residual networks, and fully convolutional neural networks into ensembles focused recognition techniques. They evaluated the system on the ISBI 2016 dataset and achieved 94.7% accuracy. They showed that the integration between different approaches could have higher performance. They performed a comparison on a fixed dataset partition, but maintaining a held-out dataset comparison is essential for a public challenge. Li and Shen [31] proposed an automated melanoma detection system using two deep learning methods. They used two fully convolutional residual networks (FCRN) simultaneously for deeper classification. They implemented the lesion feature network for dermoscopic feature extraction. They evaluated their model on ISIC 2017 and achieved an accuracy of 0.833. Kawahara et al. [32] employed pooling over a convolutional neural network for augmented feature space. They trained convolutional neural networks (CNN) using natural images to generalize classification to 10 non-dermoscopic classes. They evaluated their model on 1300 images with the 10-class dataset and achieved an accuracy of 81.8%. They focused the comparison on a partition of classes, with poorer comparison significance. Ballerini et al. [21] introduced a hierarchical k-nearest neighbor (KNN) classifier in which images were classified hierarchically into one of the two groups by the top-level classifier using low-level features. Then, within the second level classifier, the images were classified into five diagnostic classes using other subsets of features. The active contour region was used for segmenting lesions. They evaluated their model on a database which comprised of 960 image lesions and achieved a 74% classification accuracy over the five classes of skin lesions. The drawback of their method is that classification mistakes at the top level are not adjusted in the second level, which is known as the "blocking" problem. The number of misclassified images in the first level unbalanced the distribution of classes. Shrestha et al. [33] proposed a system that can discriminate malignant melanoma along with benign dysplastic nevi, using texture measures. Lesions are marked with the aid of a dermatologist. They evaluated their results on 106 dermoscopy images, 28 for melanomas, and 78 for benign dysplastic nevi and achieved an average accuracy of 95.4%. Their work focused on detecting pigment network irregularity for only the earlier melanoma, neglecting other stages in melanoma growth. Ganster et al. [34] used features of size and shape, color, and local parameters to resemble the clinical ABCD that represents the border structure, variate color, asymmetry, and dermoscopic structures defined by a dermatologist. The feature set was optimized to capture most of the significant information and then were fed into the k-nearest neighbors (KNN) classifier. They evaluated their results on 5393 skin lesion images categorized into three classes with an overall classification accuracy of 88%. It was noticed on their model that the higher cardinality subsets

had large variability in the classification performance, which resulted from the classifier over-fitting to the training data. Rezvantlab et al. [35] used CNNs in the classification of different skin diseases, using 120 images from the PH2 dataset and 10,015 from HAM10000. They achieved an average accuracy of 87.13% among skin lesion diseases. Hekler et al. [36] used CNN in the classification of skin lesion images into five diagnostic categories. They evaluated their method on 300 test images (60 for each of the five disease classes from the HAM10000 Dataset and achieved an accuracy of 82.95%. Adekanmi and Sellami [37] performed pixel-wise classification using a softmax classifier for melanoma lesions. They categorized outputs into melanoma and non-melanoma based on results derived from pixel-wise classification. They evaluated the results on the PH2 dataset and achieved 95% accuracy and 92% dice coefficient. Lynn and War [23] demonstrated a detection system for lesion borders capable of extracting relevant features of dermoscopic structures for melanoma. They used a bagging decision tree ensemble classifier to classify features extracted using the ABCD rule. The system performance was evaluated on ISBI2016, ISIC2017, and PH2 benchmarking datasets and achieved an average accuracy of 84.5%. Phillips et al. [38] proposed a Deep Ensemble for Recognition of Melanoma (DERM). Their technique was developed to identify melanoma from pigmented lesion-associated features. Their deep framework adopted a binary classification framework for recognizing melanoma from benign pigmented lesions. They trained their model using 7102 dermoscopic images, including melanoma and benign pigmented lesions. They achieved an average area under the curve (AUC) of 0.93 and an average sensitivity of 85%. Phillips et al. [39] developed an algorithm to assess suspicious from benign skin lesions. They employed a deep ensemble method for recognizing melanoma. They used 1550 images, including suspicious and benign skin lesions. They analyzed biopsied and non-biopsied pigmented skin lesions. They achieved an average AUC of 90.1% for biopsied lesions and 95.8% for other lesions. Haenssle et al. [40] proposed a CNN model to detect melanoma. They used a pre-trained Google Inception v4 model. They utilized sensitivity, specificity, and AUC for evaluating their model. They compared their CNN model against an international group of 58 dermatologists. They achieved an average specificity of 82.5%, sensitivity of 86.6%, and AUC of 88.9%. From the above-mentioned techniques, the techniques that evaluated their work on the PH2 dataset or using multiple lesion diseases were selected for performance comparison. Most of the mentioned techniques rely on capturing diagnosing feature space using low-level features that were not designed with the intent of considering the human-observable phenomenon. A feature set that contains high-level features can provide understandable justification for the system's diagnostic decisions. The pixel-based feature extraction and segmentation technique can capture a single representation that enables visualization of image structures [11]. The distributions of texture and color features enable an excellent differentiation between pigmented skin lesions from unaffected skin regions in the image [41]. Table 1 lists a summary of the comparison of the current discussed related work. For this reason, this work adopted a high-level pixel-based characterization technique for diagnosing skin image lesions to enhance diagnostic capability results. The high-level pixel-based features model proved to be a reliable biomarker for diagnosing melanocytic neoplasms progression.

Table 1. A comparison of the current related work.

Study	Image Analysis	Dataset	Methodology	Performance
Rezvantlab et al. [35]	Classification of 8 diagnostic categories of skin diseases	120 images PH2 dataset and 10,015 HAM10000	4 deep convolutional neural networks	Average Accuracy 87.13%
Hekler et al. [36]	Classification of skin lesions images into five diagnostic categories	300 images HAM10000 (60 for each disease class)	Convolutional neural networks (CNN) into a binary output	Accuracy 82.95%
Adekanmi and Viriri [37]	Classification of melanoma lesions	PH2 dataset	Softmax classifier for pixel-wise	Accuracy 95% (Binary output classifier)

Table 1. Cont.

Study	Image Analysis	Dataset	Methodology	Performance
Lynn and War [23]	Skin lesion border detection system	ISBI2016, ISIC2017, PH2 datasets	Asymmetry, border, color, and diameter (ABCD) feature extraction rule and bagging decision tree ensemble classifier	Average accuracy 84.5%
Gutman et al. [26]	Automatic detection for globules and streaks	807 training and 335 testing form ISIC 2016 dataset	Superpixels feature extraction mask in addition to dermoscopic features	Accuracy 91%
Abbadi and Faisal [22]	Detecting and segmenting malignant and benign lesion image	220 images: 120 from PH2 and 100 from the websites	YUV color space conversion, ABCD rules, segmentation based on Markov and Laplace filter	Accuracy 95.45% (Binary output classifier)
Lee et al. [28]	WonDerM pipeline (pre-processing, segmentation and classification)	HAM10000 Dataset	(DenseNet and U-net)	Accuracy 89.9%
Do et al. [27]	Detection of melanoma	117 benign nevi and 67 malignant melanomas	GLCM features, hierarchical segmentation and SVM classifier	Specificity 90%
Nammalwar et al. [29]	Segmentation of skin lesions	18 images	ABCD's clinical features, Modified Kolmogorov–Smirnov (MKS), boundary refinement algorithm	Figure comparisons
Codella et al. [30]	Segmentation and classification of melanoma	ISBI 2016 dataset	U-Net architecture, six color channels, RGB (red, green, and blue) and HSV (hue, saturation, and value) color spaces	Accuracy 76%
Li and Shen [31]	Segmentation, feature extraction, and lesion classification using two deep learning methods	ISBI 2016 dataset 2000 images	two fully convolutional residual networks (FCRN), lesion index calculation unit (LICU)	Accuracy 83.3%
Kawahara et al. [23]	Linear classifier with no lesion segmentations nor pre-processing	1300 images	multi-scale features using CNN	Accuracy 81.8%
Ballerini et al. [21]	Hierarchical classifier	960 images	k-nearest neighbors, region-based active contour	Accuracy 74%
Shrestha et al. [33]	Discrimination of early malignant melanoma	106 images	Haralick statistical texture measures	Accuracy 95.4%
Ganster et al. [34]	Early recognition of melanoma	5393 images	shape, color and local features, k-nearest neighbor (KNN) classifier	Accuracy 88%
Phillips et al. [38]	Recognition of Melanoma	7102 dermoscopic images	Deep Ensemble Model	Area under the curve (AUC) of 0.93, 85% for sensitivity
Phillips et al. [39]	Assessment of the suspicious from benign skin lesions	1550 images	artificial intelligence algorithm	AUC of 90.1% for biopsied lesions and 95.8% for other lesions
Haenssle et al. [40]	Detection of melanoma	Compared to an international group of 58 dermatologists.	CNN	82.5%, sensitivities of 86.6% and 88.9% AUC

3. The Proposed Framework

In this section, the steps of staging melanocytic neoplasms using high-level pixel-based features are discussed in detail. The proposed non-invasive staging framework consists of the following steps, as shown in Figure 6.

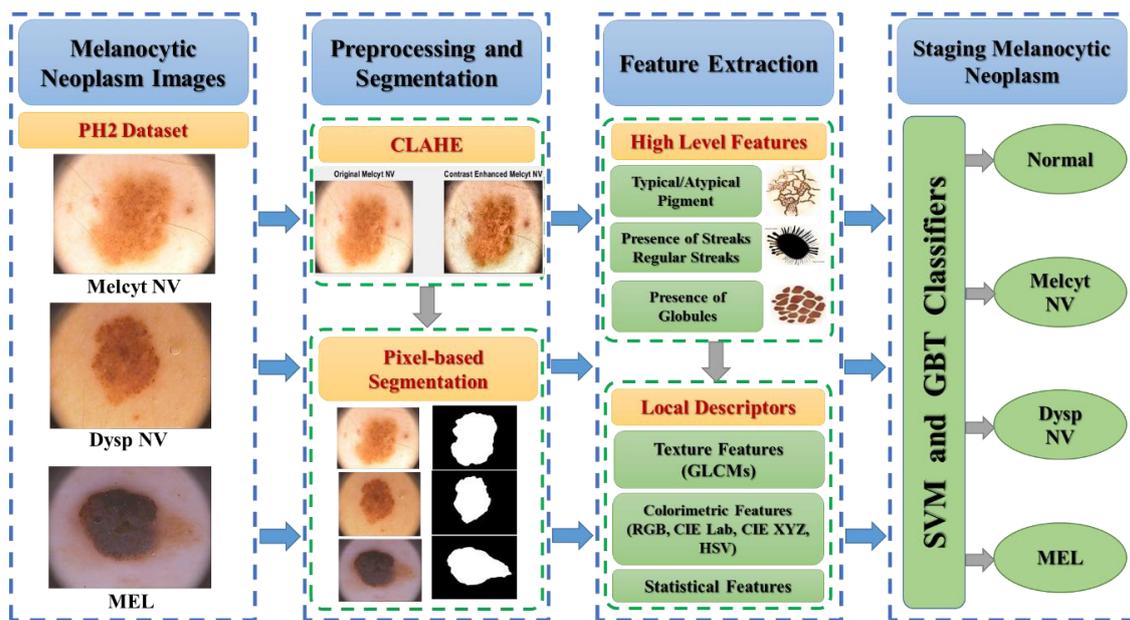


Figure 6. The non-invasive melanocytic neoplasm staging framework.

3.1. Pre-Processing and Segmentation

The melanocytic neoplasms images were enhanced using (CLAHE) contrast limited adaptive histogram equalization method. CLAHE is an effective pre-processing technique that yields better discrimination and visualization of skin image lesions [42]. CLAHE uses thresholding, equalization, and bilinear interpolation, which result in a limited homogeneous contrast region [43]. The images of the dataset have various sizes. Features extracted from different sizes of images would not have the same feature values. Thus, image pixel values were rescaled to 768×576 . The size of 768×576 was the optimal size that keeps the original image information. Thus, the images were ready for the segmentation step.

For melanocytic dermoscopic images, pixel-based segmentation was carried out in a single-pixel representation manner. The pixels of the original color images are matched with the pixels of the corresponding ground truth (GT) according to a label given by the metadata file. This technique relies on intensity levels within RGB color channel integer values and the background level. Melanocytic neoplasms were identified by intensity levels within the RGB color channel, represented by integer values, where 4 represents MEL, 2 represents the Dysp NV, 1 represents Melcyt NV, and 0 indicates the background objects. Thus, the final labels resulting from segmenting dermoscopic images had the values {0; 1; 2; 3} that can be used in staging melanocytic neoplasms. The resulting labels were used by the classifier to allocate pixel values of different melanocytic neoplasms to their corresponding stages. Figure 7 shows some examples of original Melcyt NV, Dysp NV, and mel dermoscopic images and their corresponding GTs.

3.2. Feature Extraction

Shape and pattern feature descriptors are a significant indicator affecting discrimination. Within the shape feature, every single pixel can be a feature descriptor in discriminating shapes. Thus, shapes and patterns may be represented as a single pixel, pixels in a line, a rectangular region of pixels, a polygon shape, or a region of pixels [16]. Texture can be computed based on global or local descriptors. Local descriptors can be described as statistical relationships among neighboring pixels in a region. Global descriptors can be described as computing pixel value relationships among image regions [15]. Local feature approaches are metrics used to identify the nearest range of features around the interest points within images [17]. Global feature approaches use uniform texture metrics to

generalize an entire object with a single vector, such as gray level co-occurrence matrix (GLCMs), Grey-level spatial dependency matrices, co-occurrence matrices, or extended SDMs. Color descriptors are computed by RGB-D (red, green, and blue with its corresponding depth image) data channels for greyscale, intensity, or RGB (red, green, and blue) color. Deep feature hierarchies start with local feature descriptors then produce high-level features in feature detection hierarchal layers, producing more in-depth representations [44]. Feature descriptors can also be dense or sparse, based on the selection of pixels. A dense descriptor uses all the pixels in a specified region or patch as a kernel sampling pattern, i.e., Scale Invariant Feature Transform (SIFT) and Speeded Up Robust Features (SURF). However, a sparse kernel concerns specific pixels, i.e., the local binary descriptor, where only the selected pixels are used, instead of using all pixels of the region [16].

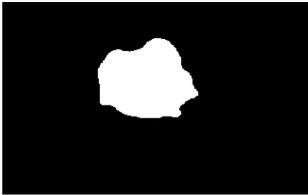
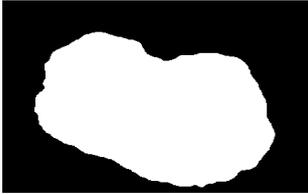
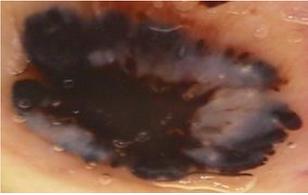
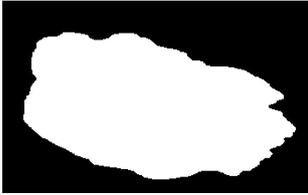
a. Original	b. Corresponding GT	High-level Features
 <p data-bbox="363 947 512 981">Melocyt NV</p>		<p data-bbox="1015 797 1315 920"> Typical pigment network Absence of Streaks Typical Globules </p>
 <p data-bbox="379 1216 496 1249">Dysp NV</p>		<p data-bbox="1015 1066 1315 1189"> Atypical pigment network Absence of Streaks Atypical Globules </p>
 <p data-bbox="411 1485 464 1518">MEL</p>		<p data-bbox="1015 1335 1315 1458"> Atypical pigment network Presence of Streaks Typical Globules </p>

Figure 7. The original Melanocytic nevi (Melocyt NV), Dysplastic nevi (Dysp NV), and malignant melanoma (Mel) dermoscopic images and their corresponding ground truths (GTs).

High-level features capture the human-observable phenomenon by describing border irregularity about a lesion image using a morphological boundary shape descriptor [45]. The incorporation between a small set of high-level features and a set of low-level features affected classification results. A morphological boundary shape descriptor is a method for processing boundary shape [20]. Morphological operations can be applied using a structuring element, such as a disk, to define the object boundary and alter the shape in some deterministic way. Morphological reshape operators clean up the shape boundary by growing or shrinking, using erode and dilate techniques [46]. Other significant indicators are also computed and incorporated into the feature space. These descriptors include color features using different color spaces, texture, and statistical features to identify regions based on texture, or statistical measures around each pixel in the input image [44].

3.2.1. Morphological Boundary Shape Descriptor

The shape descriptor concerns measuring several shapes of the pixel regions required for descriptor computations [19]. The morphological boundary shape descriptor is a method for defining polygon and boundary shape [45]. Morphological shapes generally use thresholding in defining object boundaries. The proposed high-level morphological boundary shape descriptors can give a more significant measure for the irregularity of the entire border [20]. Morphological reshape operators can clean up the lesion borders by growing or shrinking, using erode and dilate techniques [46].

Binary Dilation

The binary dilation of A by B , denoted $A \oplus B$, is defined as the set operation [47]: $A \oplus B = \{ \ddagger | (\hat{B})_{\ddagger} \cap A \neq \emptyset \}$, where \hat{B} is the structuring element B reflection [17]. The binary dilation cab is defined as the intersection between the set of reflected structuring pixel elements when translated to \ddagger , and the foreground pixels in A [46].

Grayscale Dilation

In the general form of grayscale dilation, the grayscale dilation of $A(x, y)$ by $B(x, y)$ is defined as [45]:

$$(A \oplus B)(x, y) = \max\{A(x - x', y - y') + B(x', y') | (x', y') \in D_B\}, \quad (1)$$

where D_B is the structuring element domain of B , and $A(x, y)$ is assumed to be $-\infty$ outside the domain of the image [46]. Grayscale dilation is performed with a flat structuring element ($B(x, y) = 0$). Grayscale dilation use local-maximum operator structuring element:

Binary Erosion

The binary erosion of A by B , denoted $A \ominus B$, is defined as [9]: $A \ominus B = \{ \ddagger | (B)_{\ddagger} \subseteq A \}$. The binary erosion cab is defined as the set of pixel locations \ddagger given by the intersection between the structuring pixel elements when translated to \ddagger , and the foreground pixels in A [17].

Grayscale Erosion

The grayscale erosion of $A(x, y)$ by $B(x, y)$ is defined as [47]:

$$(A \ominus B)(x, y) = \min\{A(x + x', y + y') - B(x', y') | (x', y') \in D_B\}, \quad (2)$$

where D_B is the structuring element domain of B , and $A(x, y)$ is assumed to be $+\infty$ outside the domain of the image. Grayscale erosion is generally performed with a flat structuring element ($B(x, y) = 0$). Grayscale erosion uses a local-minimum operator structuring element. The operations of morphological dilation and erosion require a flat structuring element as a binary neighborhood, two-dimensional or multidimensional. The origin of the pixels identifies the pixel in the image being processed. The opening operation is computed using a 2D-shaped structuring element, while the morphological closing operation is used to fill the gaps in an image [17].

3.2.2. Local Feature Descriptors

Local texture pixel details are computed using the statistical relationships within pixels neighborhood colorimetric in different color spaces [44]. Texture features are captured using a statistical randomness measure that assigns each output pixel to corresponding useful statistic values. Colorimetric using different color spaces, i.e., RGB, CIE XYZ, CIE Lab, and HSV for better characterization. Color channel (colorRGB) intensities are also normalized to eliminate the noise resulting from lights and shadows for better classification results. The HSV color space (colorHSV) can handle the differences in skin image lesions by removing the effect of illumination changes. Each

component can provide different information: the hue (H) component can provide intensity value without illumination changes and the component of saturation (S) can also provide a higher contrast to the image being processed. Furthermore, the CIE color spaces (labCIE and xyzCIE) can achieve higher accuracy results. Thus, they can handle the higher similarity within the skin color images [16].

3.3. Feature Reduction

The construction of feature space that is composed of high and local features resulted in a large dimensionality of features. The pixel-based feature extraction technique resulted in the generation of a large feature vector for each pixel. The feature vectors are three variables for each color space (RGB, CIE XYZ, CIE Lab, and HSV), nine variables for texture features, four descriptors for the shape, and five pigment features for the presence of globules and streaks. This ends with a total of 26 variables. To handle the curse of dimensionality, principal component analysis (PCA) is used. PCA is a projection-based technique, which projects data into a set of orthogonal axes to obtain the relevant information and dispose of the rest of the data [48]. The dimensionality reduction for large feature vectors reduced the variables to 17 significant variables. The data being transformed from high to low dimensional space affected the difficulty in processing data.

3.4. Classification of Melanocytic Dermoscopic Images

The high and local reduced feature arrays are concatenated with their corresponding labels that refer to the neoplasm phase. As mentioned previously, this work aims to stage melanocytic neoplasms dermoscopic images. The stages of melanocytic neoplasms can be Melcyt NV, Dysp NV, or Mel [23]. The SVM and Gradient Boosted Trees (GBT) classifiers are trained using reduced high-low feature space to categorize the processed pixels into neoplasms stages.

3.4.1. Support Vector Machine Classifier

SVM is a machine learning algorithm for classification. The SVM main objective is to find a dependency description between a set of object measurements, measured variables, and specific properties of these variables. Estimating the dependency between these observations can help in classifying new sets based on heuristics. Estimating the function mapping $f: \mathbb{R}^N \rightarrow \{\pm 1\}$ can determine the corresponding values for new observations [49].

Non-Separable Case

Using hard SVM, the hyperplane cannot separate in all cases. In these cases, the promotion of soft margin SVMs is needed. The soft margin SVMs can found separation hyperplanes using positive slack variables ξ_i which can be used to adapt the constraints in the following equations [50]:

$$\forall i \begin{cases} w \times x_i + b \geq +1 - \xi_i & y = +1 \\ w \times x_i + b \leq -1 - \xi_i & y = -1 \\ \xi_i \geq 0 \end{cases} \quad (3)$$

where the model parameters are the weights w , bias b , and the feature input x . Slack variables adaptation gives SVM the flexibility to reduce the optimization influence by allowing some cases to lie inside the margin or within the cases of the other class, as shown in Figure 8 [50].

RBF SVMs

The RBF kernel can handle the nonlinear cases by mapping samples non-linearly into a higher-dimensional space. The RBF utilizes two parameters: c and γ . Parameter setting issues are to identify suitable values for (c, γ) , through which the model can predict unknown data accurately. The parameter setting is commonly performed by approximations or heuristics within iterative

processes for the pair values of (c, γ) . The RBF kernel in 2D space for two inputs, x and z , can be given by Equation (4) [49].

$$k(x, z) \cong a + c(2\gamma x^T z) + q(2\gamma x^T z)^2, \tag{4}$$

The term $q(2\gamma x^T z)^2$ gives constant, linear, and second-order terms, while the value of the term $c(2\gamma x^T z)$ is bounded within the range $[0, 2\gamma]$, where $0 \leq x^T z \leq 1$ in L2-norm normalization.

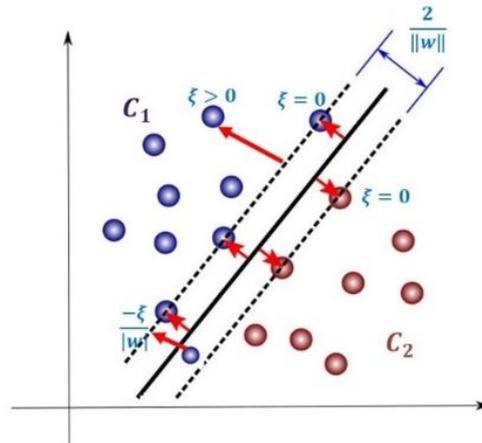


Figure 8. The soft margin support vector machines (SVM).

3.4.2. Gradient Boosted Trees:

GBT is a combination of accurate weak learners (e.g., decision trees), to create a powerful and effective predictive model [51]. The GBT ensemble boosting relies on minimizing the loss function by calculating the mean squared error between the target and predicted outputs when doing gradient boosting [52]. GBT has great popularity due to its accurateness, flexibility, and robustness, with less computational resources [52]. This is given a training set, $\{(x_1, y_1), \dots, (x_n, y_n)\}$, to minimize the expected value of loss function $L(y, F(x))$ [53]:

$$\hat{F} = \underset{F}{\operatorname{argmin}} E_{x,y}[L(y, F(x))]. \tag{5}$$

Gradient boosting approximates \hat{F} as a weighted sum of weak learners of function $h(x)$ and the coefficient γ_i [52].

$$F(x) = \sum_{i=1}^M \gamma_i h_i(x) + \text{const}. \tag{6}$$

When using GBT to fit the model, the input space is partitioned into disjoint regions $R_{1m, \dots, j}$. Thus, the base learner $h(x)$ for the tree ensemble model can be calculated by Equation (7) [53].

$$h_m(x) = \sum_{j=1}^J b_{jm} I(x \in R_{jm}), \tag{7}$$

where b_{jm} is the predicted value within R_{jm} and I is a convex loss function [54]. For each iteration, the model is updated, as in Equation (8) [52]. The update rule is modified by Friedman by using γ_{jm} for each region, instead of b_{jm} , as shown in Equation (9) [53]. The final gradient boosting can be obtained by minimizing the objectives as in Equation (10) [54].

$$F_m(x) = F_{m-1}(x) + h_m, \tag{8}$$

$$F_m(x) = F_{m-1}(x) + \sum_{j=1}^J \gamma_{jm} h_m(x) I(x \in R_{jm}), \gamma_{jm}, \tag{9}$$

$$F_m(x) = \underset{\gamma}{\operatorname{argmin}} \sum_{x_i \in R_{jm}} L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i)). \quad (10)$$

3.5. Staging Melanocytic Neoplasms

The single-pixel representation was obtained from the segmentation step, using the original color images and their corresponding GTs, resulting in identified intensity levels within the RGB color channel in integer values, where 4 represents Mel, 2 represents the Dysp NV, 1 represents Melcyt NV, and 0 indicates the background objects. Thus, the final labels resulted from segmenting dermoscopic images {0; 1; 2; 3} are used to allocate pixel values of different melanocytic neoplasms to their corresponding stages.

Label {1} Melcyt NV has a typical pigment network with uniform brown lines and regular equidistant rete ridges. Label {2} Dysp NV is characterized by a typical pigment network with irregular grid lines, globules, and streaks that may be found with the symmetrical distribution. Label {3} Mel is characterized by a typical pigment network with irregular grid lines and asymmetrical distribution of globules and streaks, with the spread of melanoma expected [55].

The formation of Mel can be seen as the deterioration of pre-malignant skin lesions in the functionality and structure of the infected lesion. They emerge from various pigmented lesions that appear on sun-damaged areas relative to limited photo-protection. Therefore, the distinction between pigmented pre-malignant and malignant lesions is a challenging task that can lead to early detection of different types of lesions and minimization of unnecessary biopsies. Melanocytic neoplasms have different dermoscopy structures, such as pigment network, globules, and streaks. Detection of streaks, globules, and pigment network is very significant in the assessment of the malignancy of a lesion [55]. Table 2 lists the pseudo-code for the staging melanocytic neoplasms.

Table 2. The pseudo-code for staging melanocytic neoplasms.

```

Start
Load the training data
Load the corresponding GT data
Load the labels CSV file
Step 1. Pre-processing:
Step 1.1 Image resize:
Images were rescaled to 768 × 576
Step 1.2 Image Enhancement:
Images were enhanced using CLAHE.
Step 1.3 Image Conversion:
Images were converted to grayscale.
Step 2. Pixel base Segmentation:
For Original Color Images
For Corresponding GT Images
For CSV labels file
Test whether a label is melcyt nv, dysp nv or mel
Use Corresponding GT mask along with Original Color Images
End for
Return single pixel labels have the values {0; 1; 2; 3}; 0 for background, 1 for melcyt nv lesion, 2 for dysp nv and
3 for mel
Save integer intensity levels array of segmented images as labels mat file.
End

```

High-level features using a morphological boundary shape descriptor can describe the border irregularity to consider the human-observable phenomenon. Melanocytic neoplasms can be Melcyt NV, Dysp NV, or Mel [45]. Abnormal growth and proliferation of abnormal cells can increase malignancy in malignant cases. Melcyt NV is characterized by a typical pigment network with uniform brown lines and regular equidistant rete ridges. Dysp NV has an atypical pigment network with irregular grid lines, globules, and streaks that may found with the symmetrical distribution. Mel characterized

by an atypical pigment network with irregular grid lines, and asymmetrical distribution of globules and streaks means that the spread of melanoma is expected. The border irregularity features, i.e., pigment network, streaks, and globules, are mapped to intuitive labels given by the PH2 dataset [56] set to construct the high-level features. The local descriptors in the form of color and texture features are also extracted in a pixel-based manner. The small set of high-level features and the set of low-level features are incorporated together to construct the feature space that is used for staging melanocytic neoplasms. The high and low-level feature space is concatenated with clinical diagnosis labeling to construct a training set to be fed into the classifier.

4. Materials and Methods

4.1. Dataset

For evaluating results, 150 skin image lesions from the PH2 dataset [56] are used. The PH2 dataset of dermoscopic images has been developed for benchmarking research purposes, to facilitate comparative studies for segmentation and classification algorithms. The database of PH2 dermoscopic images has been acquired at the Dermatology Service of Hospital Pedro Hispano, Matosinhos, Portugal. The dermoscopic images of the PH2 database are 8-bit RGB color images with a resolution of 768×560 pixels, the same conditions as the Tuebinger Mole Analyzer system, and a magnification of $20\times$. This dermoscopic database has a total of 200 dermoscopic images of melanocytic neoplasms, including 80 benign nevi (non-melanoma), 80 atypical nevi, and 40 melanomas. This database is comprised of training images and their corresponding GTs, clinical label diagnoses, and high-level intuitive labels. The PH2 dermoscopic images were rescaled to 768×576 to unify the size of the skin lesion. Different evaluation metrics were adopted to check the significance of the proposed melanoma characterization framework.

4.2. Hardware and Software Specifications

This work was implemented on an HP (Hewlett-Packard, Palo Alto, CA, USA) Envy laptop with AMD FX-7500 Radeon R7 CPU, 10 Compute Cores 4C + 6G at 2.10 GHz, and 6 GB RAM: Windows 10, 64-bit operating system, 64-based processor system type. The first processes of the algorithm were implemented by MATLAB 2018a. Then, comparisons are evaluated using RapidMiner Studio.

4.3. Performance Evaluation

The four outcomes of P positive instances and N negative instances formulates the 2×2 confusion matrix for the experiment. The area under the receiver operating characteristics (ROC) curve (AUC) is an important evaluation metric for checking the performance of the classification model. The higher the AUC, the better the model in distinguishing between classes [57]. The ROC curve is plotted with the true positive rate (TPR) against the false positive rate (FPR), where TPR is on the y -axis, and FPR is on the x -axis [57]. The accuracy (ACC) measure is used to check the capability of the classification model. Accuracy can be calculated using Equation (11). The sensitivity (Sen) or Recall measure is used to check the capability of a classifier to recognize the positive class patterns. The sensitivity of the classifier can be determined using Equation (12). The specificity (Spec) measure is used to check the capability of a classifier to recognize the patterns of negative class. It can be calculated using Equation (13). The F measure or dice similarity coefficient (DSC) considers both precision and recall, measuring the accuracy of the test. DSC ranges from 0 worst score to 1 best score as a weighted average of precision and recall. The DSC measure can be calculated using Equation (14) [24].

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \times 100, \quad (11)$$

$$\text{Sen} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (12)$$

$$\text{Spec} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \quad (13)$$

$$\text{DSC} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (14)$$

4.4. Results

Various classification models are used for comparisons to evaluate the proposed technique. SVM, GBT, random forest (RF), Naïve Bayes (NB), and deep learning (DL) classifiers are used for comparison. Significant evaluation metrics are used to check the capability of the distinguishing between classes. ACC, AUC, DSC, Sen, and Spec are used as performance indicators. The proposed framework achieved higher performance results using the SVM and GBT classifiers. During the training of the GBT classifier, the optimal parameters were 150 trees in the forest with a maximal depth of 7. The optimal learning rate was 0.1. The RBF kernel parameters were exploited in training, gamma was set to 0.01, and the C parameter was set to 1000. The SVM model included a total of 1176 support vectors and 1.234 Bias (offset).

To evaluate the results of the proposed system, the computed results were compared with other different state-of-the-art classifiers, which are RF, NB, and DL. RF classifier utilized 140 decision trees in the forest, with a maximal depth of 7. The NB model built the Bayesian classification method. The data distribution was modeled with best-fit Gaussian and multinomial distribution. The DL model is constructed based on a multi-layer feed-forward neural network, which is based on back-propagation. This model was used as a classifier, which was trained by the previously extracted feature. The network structure consists of an input layer, three hidden layers, and an output layer. The input layer consists of 17 input feature neurons resulting from dimensionality reduction. The first and second hidden layers consist of 50 neurons each. The last hidden layer consists of 25 neurons. The output layer has four neurons based on the number of tested classes. We conducted hyper-parameterization to choose the optimal values for the learning rate, momentum training, annealing rate, regularization, and loss function in order to enable high predictive accuracy. The number of images was extended using an augmentation technique in different transformations, i.e., rotation, shifting, scaling (zoom in/out), and flipping. The parameters and their optimal values for the comparing classifiers are listed in Table 3.

Table 3. The parameters of the comparison classifiers.

Classifier	Parameters	Value
SVM	Kernel	RBF
	Gamma	0.01
	C parameter	1000
	Support vectors	1176
	Bias (offset)	1.234
GBT	No. of trees	150
	Maximal depth	7
	Learning rate	0.1
RF	No. of decision trees	140
	Maximal depth	7
NB	NB model	Bayesian
	Distribution best-fit	Gaussian
	Distribution type	multinomial
DL	No. of epochs	20
	Activation function	ReLU
	Loss function	Cross-Entropy
	ϵ	1.0×10^{-8}
	L1	1.0×10^{-5}
	L2	0

The 10-fold cross-validation technique is employed to evaluate the performance of the proposed system. For the 10-fold cross-validation, the dataset is split into 80 for training and 20 for the validation set. The performance evaluation using 10-fold cross-validation for the proposed technique against other classifiers can be seen in Table 4.

Table 4. Ten-fold cross-validation performance evaluation.

Model	ACC	AUC	DSC	Sen	Spec	Recall	Precision	Total Time
SVM	92.2% ± 0.3	0.969 ± 0.012	95.1% ± 0.2	99.0% ± 0.2	69.9% ± 5.3	99.0% ± 0.2	91.5% ± 0.3	25 min, 55 s
GBT	92.1% ± 0.1	0.962 ± 0.002	95.0% ± 0.1	98.5% ± 0.1	71.1% ± 0.8	98.5% ± 0.1	91.8% ± 0.2	12 min, 40 s
RF	90.0% ± 0.2	0.945 ± 0.003	93.6% ± 0.1	95.4% ± 0.3	72.5% ± 0.9	95.4% ± 0.3	91.9% ± 0.2	20 min, 53 s
NB	84.7% ± 0.0	0.906 ± 0.0	89.9% ± 0.0	89.1% ± 0.0	70.1% ± 0.2	89.1% ± 0.0	90.7% ± 0.0	6 min, 1 s
DL	84.1% ± 0.1	0.898 ± 0.0	90.5% ± 0.1	98.9% ± 0.0	35.8% ± 0.5	98.9% ± 0.0	83.5% ± 0.1	8 min, 39 s

Regarding the comparison measurements in Table 4, the higher performance was achieved using the SVM classifier, rather than GBT. The proposed system achieved an average ACC of 92.2%, AUC of 0.969, DSC of 95.1%, Sen of 99.0%, and Spec of 69.9% using SVM. Besides, the proposed system achieved an average ACC of 92.1%, AUC of 0.962, DSC of 95.0%, Sen of 98.5% and Spec of 71.1% using GBT. RF, NB, and DL achieved average ACCs of 90.0%, 84.7%, and 84.1%, respectively. They achieved average AUCs of 0.945, 0.906, and 0.898, respectively. They achieved average DSCs of 93.6%, 89.9%, and 90.5%, respectively. They achieved average Sens of 95.4%, 89.1%, and 98.9%, respectively. They achieved average Specs of 72.5%, 70.1%, and 35.8%, respectively. The results of the proposed system, using the SVM and GBT classifiers, outperforms other state-of-the-art techniques. The SVM model achieved higher results because of its potential for high accuracy with few training sets [24]. The SVM multiclass classifier has the ability to map the class of interest, locate the support vectors, and use the optimal kernel function that makes the classifier more flexible and robust against the outliers [25].

In addition, we used a four-fold cross-validation technique to validate the obtained results. For the four-fold cross-validation, the data set is split into 70 for training and 40 for the validation set. Table 5 shows a comparison of the proposed technique against other classifiers by using a four-fold cross-validation technique. Higher performance was achieved using the SVM classifier than GBT. The proposed system achieved an average ACC of 92.9%, AUC of 0.959, DSC of 95.3%, Sen of 98.8%, and Spec of 86.7% using SVM. Furthermore, the proposed system achieved an average ACC of 92.6%, AUC of 0.959, DSC of 94.2%, Sen of 93.5%, and Spec of 77.5% using GBT.

Table 5. Four-fold cross-validation performance evaluation.

Model	ACC	AUC	DSC	Sen	Spec	Recall	Precision	Total Time
SVM	92.9% ± 0.3	0.959 ± 0.007	95.3% ± 0.2	98.8% ± 0.1	86.7% ± 1.5	98.8% ± 0.1	94.8% ± 0.6	16 min, 22 s
GBT	91.6% ± 0.4	0.959 ± 0.001	94.2% ± 0.3	93.5% ± 0.9	77.5% ± 0.6	93.5% ± 0.9	92.0% ± 0.5	11 min, 14 s
RF	89.8% ± 0.4	0.945 ± 0.002	92.9% ± 0.2	92.6% ± 0.3	82.6% ± 1.0	92.6% ± 0.3	93.3% ± 0.4	15 min, 15 s
NB	85.2% ± 0.1	0.923 ± 0.0	89.2% ± 0.0	84.8% ± 0.0	86.4% ± 0.2	84.8% ± 0.1	94.2% ± 0.1	4 min, 13 s
DL	90.1% ± 0.1	0.956 ± 0.002	93.6% ± 0.0	99.9% ± 0.0	64.8% ± 0.3	99.9% ± 0.0	88.0% ± 0.1	7 min, 35 s

The comparison measurements for other classifiers, RF, NB, and DL, achieved average ACCs of 89.8%, 85.2%, and 90.1%, respectively. They achieved an average AUCs of 0.945, 0.923, and 0.956, respectively. They achieved an average DSCs of 92.9%, 89.2%, and 93.6%, respectively. They achieved an average Sens of 92.6%, 86.4%, and 64.9%, respectively. They achieved an average Specs of 72.5%, 70.1%, and 35.8%, respectively. The results of the proposed system, using SVM and GBT classifiers, outperforms other state-of-the-art techniques.

To visualize the performance for the proposed diagnostic system, a ROC curve was constructed for the proposed model, along with the other tested classifier. A ROC curve is created by plotting the TPR against the FPR. Figure 9 shows the relationship between sensitivity and specificity for all tested classifiers. The results show the quality of the proposed model's predictions, along with the other tested classifiers.

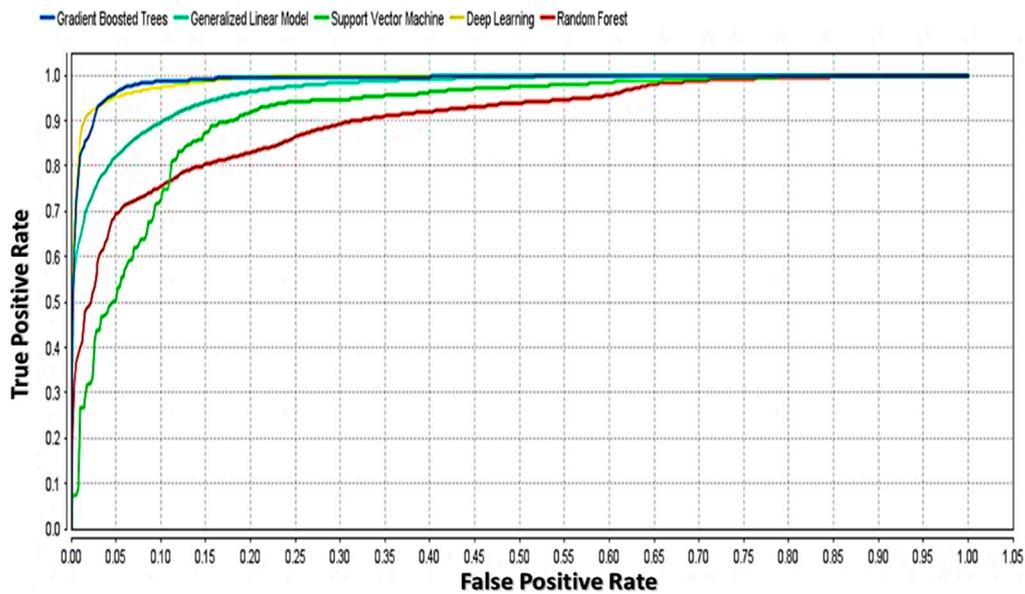


Figure 9. The area under the receiver operating characteristics (ROC) curve for the tested classifiers.

The literature works include some techniques based on a binary output classifier and others based on a multiclass output classifier for evaluating their results. The multiclass output techniques were used for performance comparison with the proposed framework. Performance comparison of researchers that evaluated their work on the same PH2 dataset and adopted a multiclass output classifier as the proposed work is presented. Performance comparison with state-of-the-art multiclass techniques is shown in Figure 10.

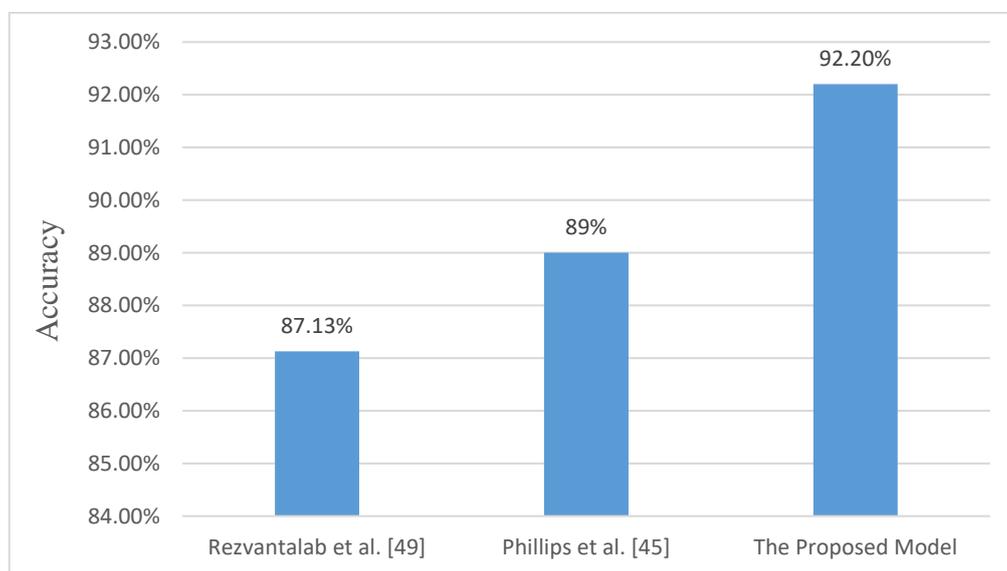


Figure 10. Performance comparison with state-of-the-art multiclass techniques based on accuracy.

4.5. Discussion

Earlier dermoscopic techniques relied on capturing diagnosing feature space using low-level features that were not designed with the intent of considering the human-observable phenomenon. A feature set containing high-level features can provide understandable justification for the system’s diagnostic decisions. The pixel-based feature extraction and segmentation technique can capture a single representation that enables the visualization of image structures. The distributions of

texture and color features enable an excellent differentiation between pigmented skin lesions from unaffected skin regions in the image. For this reason, this work adopted a high-level pixel-based characterization technique for diagnosing skin image lesions to enhance diagnostic capability results. The proposed model is evaluated on the PH2 dataset of dermoscopic images acquired from Hospital Pedro Hispano, Portugal. Several researchers have evaluated their model on the PH2 dataset of dermoscopic images. Some of these researchers evaluated their model on the PH2 dataset, i.e., Adekanmi et al. [37], Lynn et al. [23], and Abbadi et al. [22], and adopted a binary output classifier in their model. Adekanmi et al. [37], Lynn et al. [23], and Abbadi et al. [22] achieved average accuracies of 95%, 84.5%, and 95.45%, respectively, but in binary output classifiers.

Other researchers, i.e., Rezvantalab et al. [35], evaluated their model on the PH2 dataset and adopted a multiclass output classifier for evaluating their results. Rezvantalab et al. [35] achieved 87.13% accuracy adopting a multiclass classifier on the PH2 dataset. Others adopted a multiclass output classifier for evaluating their model using a different dataset, i.e., Hekler et al. [36] and Codella et al. [30]. Hekler et al. [36] and Codella et al. [30] achieved accuracies of 82.95% and 76%, respectively. The proposed technique outperforms the results achieved by other researchers, achieving 92.2% multiclass classification on the PH2 dataset for characterizing different melanocytic neoplasms stages.

5. Conclusions

This paper proposes a comprehensive pixel-based framework for staging melanocyte neoplasms. The framework proposed uses high-level analytical reasoning to describe border irregularity, in addition to various feature descriptors, i.e., color, texture. Different types of features are derived from staging growing of lesions, from benign lesions in terms of NV up to pre-malignant Dysp NV or malignant melanoma. The distributions of texture and color features enable differentiation between pigmented and unaffected skin regions within the images. The adopted a high-level pixel-based technique assisted the extraction of significant features for training the model. These features are color features in different color spaces, local statistics, and texture morphology. The mapping between the high-level features to intuitive labels given by the PH2 data set assisted the construction of feature space. The incorporation between a small set of high-level and low-level features affected the classification results. Staging melanocyte neoplasms were carried out by training SVM and GBT classifiers, with the extracted feature space along with clinical labels. The results show that the proposed system can help in guiding the diagnosis of pigmented skin lesions at different stages. The diagnosis of skin lesions at earlier stages can help in improving the durability of skin cancer and reducing the skin cancer mortality rate. For future works, more analysis of the images and expansion of the image database is required for more promising results. On the other hand, DL performs better with a large number of data. Therefore, future works will also seek to investigate DL with a large skin cancer dataset. We will work to employ a CNN with a multi-path to classify different grades of skin cancer.

Author Contributions: Conceptualization, M.R.I. and M.E.; methodology, M.R.I. and M.E.; software, M.R.I. and M.E.; validation, S.E.-S. and T.A.; formal analysis, M.R.I. and M.E.; investigation, M.R.I., S.E.-S., T.A., and M.E.; resources, S.E.-S. and T.A.; data curation, M.R.I. and M.E.; writing—original draft preparation, M.R.I. and M.E.; writing—review and editing, M.R.I., S.E.-S., T.A., and M.E.; visualization, M.R.I., T.A., and M.E.; supervision, M.E.; project administration, M.E.; funding acquisition, S.E.-S. and T.A. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (NRF-2016R1D1A1A03934816).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tschandl, P.; Wiesner, T. Advances in the diagnosis of pigmented skin lesions. *Br. J. Dermatol.* **2018**, *178*, 9–11. [[CrossRef](#)] [[PubMed](#)]

2. Russo, T.; Piccolo, V.; Ferrara, G.; Agozzino, M.; Alfano, R.; Longo, C.; Argenziano, G. Dermoscopy pathology correlation in melanoma. *J. Dermatol.* **2017**, *44*, 507–514. [[CrossRef](#)] [[PubMed](#)]
3. Jeffrey, G.; Danielle, S.; Orenge, I. Common Adult Skin and Soft Tissue Lesions. *Semin Plast Surg.* **2016**, *30*, 98–107.
4. Ankad, B.S.; Sakhare, P.S.; Prabhu, M.H. Dermoscopy of non-melanocytic and pink tumors in brown skin: A descriptive study. *Dermatopathol. Diagn. Dermatol.* **2017**, *4*, 41–51. [[CrossRef](#)]
5. Damsky, W.; Bosenberg, M. Melanocytic nevi and melanoma: Unraveling a complex relationship. *HHS Public Access* **2017**, *36*, 5771–5792. [[CrossRef](#)]
6. Jason, P.; Lott, M. *Almost one in Four Skin Biopsies is Melanocytic Proliferation*; Medical Press: New Haven, CT, USA, 2017.
7. Cannavò, S.P.; Tonacci, A.; Bertino, L.; Casciaro, M.; Borgia, F.; Gangemi, S.; Casciaro, M.; Borgia, F. The role of oxidative stress in the biology of melanoma: A systematic review. *Pathol. Res. Pract.* **2019**, *215*, 21–28. [[CrossRef](#)]
8. Abbasi, N.R.; Shaw, H.M.; Rigel, D.S.; Friedman, R.J.; McCarthy, W.H.; Osman, I.; Kopf, A.W.; Polsky, D. Early Diagnosis of Cutaneous Melanoma. *JAMA* **2004**, *292*, 2771–2776. [[CrossRef](#)]
9. Philip, C. Benign pigmented skin lesions. *AJGP* **2019**, *48*, 364–367.
10. Kittler, H.; Marghoob, A.A.; Argenziano, G.; Carrera, C.; Curiel-Lewandrowski, C.; Hofmann-Wellenhof, R.; Malvehy, J.; Menzies, S.; Puig, S.; Rabinovitz, H.; et al. Standardization of terminology in dermoscopy/dermatoscopy: Results of the third consensus conference of the International Society of Dermoscopy. *J. Am. Acad. Dermatol.* **2016**, *74*, 1093–1106. [[CrossRef](#)]
11. Khalil, A.; Elmogy, M.; Ghazal, M.; Burns, C.; El-Baz, A. Chronic Wound Healing Assessment System Based on Different Features Modalities and Non-Negative Matrix Factorization (NMF) Feature Reduction. *IEEE Access* **2019**, *7*, 80110–80121. [[CrossRef](#)]
12. Anantha, M.; Moss, R.; Stoecker, W. Detection of pigment network in dermatoscopy images using texture analysis. *Comput Med. Imaging Graph.* **2011**, *28*, 225–234. [[CrossRef](#)] [[PubMed](#)]
13. Abbes, W.; Sellami, D. Automatic Skin Lesions Classification Using Ontology-Based Semantic Analysis of Optical Standard Images. *Procedia Comput. Sci.* **2017**, *112*, 2096–2105. [[CrossRef](#)]
14. Zaqout, I. Diagnosis of Skin Lesions Based on Dermoscopic Images Using Image Processing Techniques. In *Pattern Recognition—Selected Methods and Applications*; Intechopen Limited: London, UK, 2019. [[CrossRef](#)]
15. Krig, S. Interest Point Detector and Feature Descriptor Survey. In *Computer Vision Metrics*; Apress: Berkeley, CA, USA, 2014.
16. Krig, S. Global and Regional Features. In *Computer Vision Metrics*; Apress: Berkeley, CA, USA, 2014.
17. Raju, S.; Rajan, E. Skin Texture Analysis Using Morphological Dilation and Erosion. *Int. J. Pure Appl. Math.* **2018**, *118*, 205–223.
18. Olugbara, O.; Taiwo, T.; Heukelman, D. Segmentation of Melanoma Skin Lesion Using Perceptual Color Difference Saliency with Morphological Analysis. *Math. Probl. Eng.* **2018**. [[CrossRef](#)]
19. Descombes, X.; Komech, S. Shape Descriptor Based on the Volume of Transformed Image Boundary. In *Pattern Recognition and Machine Intelligence*; Springer: Berlin, Germany, 2011; pp. 142–147.
20. Amelard, R.; Wong, A.; Clausi, D. Extracting morphological high-level intuitive features (HLIF) for enhancing skin lesion classification. *Conf Proc. IEEE Eng. Med. Biol Soc.* **2012**, *2012*, 4458–4461.
21. Ballerini, L.; Fisher, R.; Aldridg, B.; Rees, J.L. A Color and Texture Based Hierarchical K-NN Approach to the Classification of Non-melanoma Skin Lesions. In *Color Medical Image Analysis*; Springer Science: Berlin, Germany, 2013.
22. Abbadi, N.; Faisal, Z. Detection and Analysis of Skin Cancer from Skin Lesions. *Int. J. Appl. Eng. Res.* **2017**, *12*, 9046–9052.
23. Lynn, N.; War, N. Melanoma Classification on Dermoscopy Skin Images using Bag Tree Ensemble Classifier. In Proceedings of the International Conference on Advanced Information Technologies (ICAIT), Yangon, Myanmar, 6–7 November 2019.
24. Ibraheem, M.R.; Elmogy, M. Automated Segmentation and Classification of Hepatocellular Carcinoma Using Fuzzy C-Means and SVM. In *Medical Imaging in Clinical Applications, Studies in Computational Intelligence*; Springer International Publishing: Cham, Switzerland, 2016.
25. Moughal, T.A. Hyperspectral image classification using Support Vector Machine. *J. Phys. Conf. Ser.* **2013**, *439*. [[CrossRef](#)]

26. Codella, N.C.F.; Gutman, D.; Celebi, M.E.; Helba, B.; Marchetti, M.A.; Dusza, S.W.; Kalloo, A.; Liopyris, K.; Mishra, N.; Kittler, H.; et al. Skin Lesion Analysis toward Melanoma Detection: A Challenge at the International Symposium on Biomedical Imaging (ISBI) 2016, hosted by the International Skin Imaging Collaboration (ISIC). *arXiv* **2016**, arXiv:1605.01397.
27. Do, T.-T.; Hoang, T.; Pomponiu, V.; Zhou, Y.; Chen, Z.; Cheung, N.-M.; Koh, D.; Tan, A.; Tan, S.-H.; Zhao, C.; et al. Accessible Melanoma Detection using Smartphones and Mobile Image Analysis. *IEEE Trans. Multimed.* **2018**, *20*, 2849–2864. [[CrossRef](#)]
28. Lee, Y.; Jung, S.; Won, H. WonDerM: Skin Lesion Classification with Fine-tuned Neural Networks. Available online: <https://arxiv.org/abs/1808.03426> (accessed on 3 September 2020).
29. Nammalwar, P.; Ghita, O.; Whelan, P. Integration of Colour and Texture Distributions for Skin Cancer Image Segmentation. Available online: https://www.researchgate.net/publication/236645646_Integration_of_Colour_and_Texture_Distributions_for_Skin_Cancer_Image_Segmentation (accessed on 3 September 2020).
30. Codella, N.C.F.; Nguyen, Q.-B.; Pankanti, S.; Gutman, D.; Helba, B.; Halpern, A.C.; Smith, J.R. Deep learning ensembles for melanoma recognition in dermoscopy images. *IBM J. Res. Dev. USA* **2017**, *6*, 4–5. [[CrossRef](#)]
31. Li, Y.; Shen, L. Skin Lesion Analysis towards Melanoma Detection Using Deep Learning Network. *Sensors* **2018**, *18*, 556. [[CrossRef](#)] [[PubMed](#)]
32. Kawahara, J.; BenTaieb, A.; Hamarneh, G. Deep features to classify skin lesions. In *International Symposium on Biomedical Imaging (ISBI)*; IEEE: Prague, Czech Republic, 2016.
33. Shrestha, B.; Bishop, J.; Kam, K.; Chen, X.; Moss, R.H.; Stoecker, W.V.; Umbaugh, S.; Stanley, R.J.; Celebi, M.E.; Marghoob, A.A.; et al. detection of atypical texture features in early malignant melanoma. *Ski. Res. Technol.* **2010**, *16*, 60–65. [[CrossRef](#)] [[PubMed](#)]
34. Ganster, H.; Pinz, A.; Röhner, R. Automated Melanoma Recognition. *IEEE Trans. ON Med Imaging* **2001**, *20*, 233–239. [[CrossRef](#)] [[PubMed](#)]
35. Rezvantalab, A.; Safigholi, H.; Karimijeshni, S. Dermatologist Level Dermoscopy Skin Cancer Classification Using Different Deep Learning Convolutional Neural Networks Algorithms. *Comput. Vis. Pattern Recognit. arXiv* **2018**, arXiv:1810.10348.
36. Hekler, A.; Utikal, J.S.; Enk, A.H.; Hauschild, A.; Weichenthal, M.; Maron, R.C.; Berking, C.; Haferkamp, S.; Klode, J.; Schadendorf, D.; et al. Superior skin cancer classification by the combination of human and artificial intelligence. *Eur. J. Cancer* **2019**, *120*, 114–121. [[CrossRef](#)]
37. Adekanmi, A.; Viriri, S. Deep Learning-Based System for Automatic Melanoma Detection. *IEEE Access* **2020**, *8*, 7160–7172.
38. Phillips, M.; Greenhalgh, J.; Marsden, H.; Palamaras, L. Detection of Malignant Melanoma Using Artificial Intelligence: An Observational Study of Diagnostic Accuracy. *Dermatol. Pract. Concept.* **2020**, *10*, e2020011.
39. Michael Phillips, M.; Marsden, H.; Jaffe, W.; Matin, R.N.; Wali, G.N.; Greenhalgh, J.; McGrath, E.; James, R.; Ladoyanni, E.; Bewley, B.; et al. Assessment of Accuracy of an Artificial Intelligence Algorithm to Detect Melanoma in Images of Skin Lesions. *JAMA Netw. Open* **2019**, *2*, e1913436. [[CrossRef](#)]
40. Haenssle, H.A.; Fink, C.; Schneiderbauer, R.; Toberer, F.; Buhl, T.; Blum, A.; Kalloo, A.; Hassen, A.B.H.; Thomas, L.; Enk, A.; et al. Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann. Oncol.* **2018**, *29*, 1836–1842. [[CrossRef](#)]
41. Verma, A.; Pal, S.; Kumarb, S. Comparison of skin disease prediction by feature selection using ensemble data mining techniques. *Inform. Med. Unlocked* **2019**, *16*, 100202. [[CrossRef](#)]
42. Hoshyar, A.N.; Al-Jumailya, A.; Hoshyar, A.N. The Beneficial Techniques in Pre-processing Step of Skin Cancer Detection System Comparing. In *Procedia Computer Science*; Elsevier: Amsterdam, The Netherlands, 2014; pp. 25–31.
43. Campos, G.; Mastelini, S.; Aguiar, G. Machine learning hyperparameter selection for Contrast Limited Adaptive Histogram Equalization. *EURASIP J. Image Video Process.* **2019**, *59*. [[CrossRef](#)]
44. Krig, S. Local Feature Design Concepts, Classification, and Learning. In *Computer Vision Metrics*; Apress: Berkeley, CA, USA, 2014.
45. Iwanowski, M. Morphological Boundary Pixel Classification. In Proceedings of the International Conference on “Computer as a Tool”, Warsaw, Poland, 9–12 September 2007.
46. Ramkumar, P. Morphological Representation Operators, Algorithms And Shape Descriptors. Available online: https://shodhganga.inflibnet.ac.in/bitstream/10603/40771/8/08_chapter3.pdf (accessed on 3 September 2020).

47. Banerjee, S.; Sahasrabudhe, S. A morphological shape descriptor. *J. Math. Imaging Vis.* **1994**, *4*, 43–55. [CrossRef]
48. Zhang, L.; Dong, W.; Zhang, D.; Shi, G. Two-stage image denoising by principal component analysis with local pixel grouping. *Pattern Recognit.* **2010**, *43*, 1531–1549. [CrossRef]
49. Cao, H.; Naito, T.; Ninomiya, Y. Approximate RBF Kernel SVM and Its Applications in Pedestrian Classification. Available online: https://www.researchgate.net/publication/29621872_Approximate_RBF_Kernel_SVM_and_Its_Applications_in_Pedestrian_Classification (accessed on 3 September 2020).
50. Afentoulis, V.; Lioufi, K. Svm Classification With Linear And Rbf Kernels. Available online: https://www.researchgate.net/publication/279913074_SVM_Classification_with_Linear_and_RBF_kernels (accessed on 3 September 2020).
51. Song, B.; Sacan, A. Automated wound identification system based on image segmentation and artificial neural networks. In Proceedings of the 2012 IEEE International Conference on Bioinformatics and Biomedicine, Philadelphia, PA, USA, 4–7 October 2012; pp. 1–4.
52. Liu, S.; Xiao, J.; Liu, J.; Wang, X.; Wu, J.; Zhu, J. Visual Diagnosis of Tree Boosting Methods. *IEEE Trans. Vis. Comput. Graph.* **2018**, *24*, 163–173. [CrossRef]
53. Yang, J. Applying Boosting Algorithm for Improving Diagnosis of Interstitial Lung Diseases. Available online: <http://cs229.stanford.edu/proj2016/report/YangApplyingBoostingAlgorithmForImprovingDiagnosisOfInterstitialLungDisease-report.pdf> (accessed on 3 September 2020).
54. Friedman, J.H. Machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
55. Jiménez, A.; Serrano, C.; Acha, B.; Karray, F.; Campilho, A.; Cheriet, F. Automatic Detection of Globules, Streaks and Pigment Network Based on Texture and Color Analysis in Dermoscopic Images. In *Bioinformatics Research and Applications*; Springer Science: Berlin, Germany, 2017; pp. 486–493.
56. PH2 Dataset. Available online: <https://www.fc.up.pt/addi/ph2%20database.html> (accessed on 3 September 2020).
57. Bradley, P.; Andrew, P. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit* **1997**, *30*, 1145–1159. [CrossRef]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).