applied sciences

MDPI

Article

An Identification Method of Feature Interpretation for Melanoma Using Machine Learning

Zhenwei Li *,†, Qing Ji †, Xiaoli Yang, Yu Zhou and Shulong Zhi

School of Medical Technology and Engineering, Henan University of Science and Technology, No. 263, Kaiyuan Avenue, Luoyang 471023, China; 210321221634@stu.haust.edu.cn (Q.J.); yxl@haust.edu.cn (X.Y.); yuzhou@haust.edu.cn (Y.Z.); 220320221779@stu.haust.edu.cn (S.Z.)

* Correspondence: lizhenwei@haust.edu.cn

† These authors contributed equally to this work.

Abstract: Melanoma is a fatal skin cancer that can be treated efficiently with early detection. There is a pressing need for dependable computer-aided diagnosis (CAD) systems to address this concern effectively. In this work, a melanoma identification method with feature interpretation was designed. The method included preprocessing, feature extraction, feature ranking, and classification. Initially, image quality was improved through preprocessing and k-means segmentation was used to identify the lesion area. The texture, color, and shape features of this region were then extracted. These features were further refined through feature recursive elimination (RFE) to optimize them for the classifiers. The classifiers, including support vector machine (SVM) with four kernels, logistic regression (LR), and Gaussian naive Bayes (GaussianNB) were applied. Additionally, cross-validation and 100 randomized experiments were designed to guarantee the generalization of the model. The experiments generated explainable feature importance rankings, and importantly, the model demonstrated robust performance across diverse datasets.

Keywords: skin cancer; melanoma; image classification; feature selection; interpretability

check for updates

Citation: Li, Z.; Ji, Q.; Yang, X.; Zhou, Y.; Zhi, S. An Identification Method of Feature Interpretation for Melanoma Using Machine Learning. *Appl. Sci.* **2023**, *13*, 10076. https://doi.org/10.3390/app131810076

Academic Editors: Mohamed Medhat Gaber and Mohammed Abdelsamea

Received: 12 August 2023
Revised: 30 August 2023
Accepted: 5 September 2023
Published: 7 September 2023

Copyright: © 2023 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 (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Skin diseases encompass a wide range of conditions, some of which can progress into severe and fatal skin cancers. The World Health Organization projects that there will be approximately 2.2 million cases of skin cancer by 2025 [1]. Among skin cancers, melanoma stands out due to its very high mortality rate; its prevalence continues to rise, with over 130,000 new cases diagnosed [2]. Research has established a strong link between exposure to sunlight and ultraviolet radiation and the development of melanoma. Prolonged exposure to ultraviolet radiation, along with cellular mutations or genetic defects, can lead to the mutation and rapid proliferation of normal melanocytes in the basal layer of the epidermis, culminating in the progression to melanoma. Due to its aggressive nature and high mutation frequency, early diagnosis of melanoma is of utmost importance. At the initial stage, the success rate of cure can exceed 90% [3].

For the accurate early diagnosis of skin diseases, dermoscopic images are often employed. Dermoscopy, a non-invasive imaging technique, provides a magnified view of the deeper layers of the skin under non-invasive conditions, facilitating the analysis and diagnosis of skin lesions and affected areas [4]. The utilization of dermoscopic images enhances medical observation, significantly improving the diagnosis of various skin diseases.

Image processing for the detection, segmentation, and classification of skin lesions encounters various difficulties [5], including: (a) the presence of noise and artifacts such as hairs, bubbles, and blood vessels; (b) irregular, random, and sometimes diffuse edges with low contrast between the lesion and healthy skin; (c) illumination malfunctions; and (d) variability in image characteristics due to different types of capturing equipment. Due to these challenges, image preprocessing is essential before diagnosing melanoma.

In recent years, the rapid advancement of artificial intelligence and machine learning technologies has garnered significant attention from researchers, leading to the development of automated image processing systems for diagnosing diseases in the medical field. Various image CAD systems have been developed for the diagnostic classification of melanoma, providing valuable assistance to physicians [6].

The CAD system of melanoma based on machine learning typically involves three steps:

1. Segmentation of the skin lesion region;
2. Feature extraction from the skin lesion region;
3. Classification of the extracted features using a classifier.

To achieve more accurate classification results, segmentation is crucial, as it provides precise samples for the classifier. While numerous studies have focused on modifying CAD systems for melanoma classification, a limited number have analyzed the features extracted from melanoma and explored their link with the pathological perspective.

Dermoscopic images can pose challenges for identification due to hair problems or acquisition processes, but these issues can be addressed through image preprocessing. Seena, J. et al. [7] demonstrated that preprocessing significantly impacts the segmentation process of skin lesion areas, leading to more accurate results. Proper segmentation is crucial as it affects the subsequent feature extraction and melanoma identification, making preprocessing necessary. Ashraf, H. et al. [8] and Bakheet, S. et al. [9] resized images and employed simple filtering to address hair effects and noise elimination.

The preprocessing stage primarily aims to improve image quality without altering the size of the skin lesion region. Segmentation can also be integrated into the preprocessing stage, utilizing traditional algorithms like adaptive threshold segmentation [10] and Otsu's clustering-based approach [11].

Automated CAD systems are commonly based on various feature descriptions, including color, texture, and shape [12]. In some studies, color and texture information from dermoscopic images were extracted for melanoma identification, followed by classification using classifiers. Rastgoo, M. et al. [13] proposed a classification framework that compared shape, texture, and color features both locally and holistically to differentiate between melanomas and dysplastic nevi. They utilized SVMs, gradient boosting (GB), and random forest (RF) classifiers, considering single and combined features.

During feature extraction, Bharathi, G. et al. [14] employed color map histogram equalization and a fuzzy system to enhance dermoscopic images, using a genetic algorithm (GA) to optimize extracted texture features for melanoma detection. Nasir, M. et al. [15] utilized the Boltzmann entropy method to select fused texture, color, and shape features, and employed SVM for classification.

Previous research in this field has concentrated on extracting diverse features like texture, color, and shape to enhance melanoma recognition methods or models. However, few studies have focused on the importance of features and the interpretability of models. For instance, a study by Wahba, M.A. [16] et al. extracted texture features using the GLCM method enhanced by the GLDM technique. However, a detailed analysis of these features was lacking. Moldovanu, S. [17] et al. predominantly extracted color features in the BGR color space, with an opportunity to explore the HSV color space for broader color conditions. Feature selection, as demonstrated by Chatterjee, S. [4] et al., employed the RFE method. However, the focus was on optimal features without a comprehensive selection study. Lin et al. [18] discussed the capacity of the RFE method to rank features. Visualizing the RFE method as a gold standard for encapsulated embedding, as seen in the work by Sanz H [19], remains limited.

The trend leans towards improving the accuracy in melanoma identification methods or models. While accuracy is crucial, model generalizability holds equal significance. Explaining the internal role of model features through visualization can increase the credibility of the model. Moreover, bestowing a degree of generalizability to the model through multifaceted design holds valuable research potential.

This paper designs a randomized experiment that employs RFE's feature importance ranking and cross-validation for melanoma identification. The goal is to enhance the credibility of the model and ensure its broader applicability. The process involves preprocessing dermoscopic images to enhance lesion regions and remove impurities. Segmentation using the k-means method extracts lesion regions, followed by cropping to isolate background effects. Three features—color, texture, and shape—are extracted from the lesion region and visualized. RFE is used to filter influential features based on the visualization insights. This approach helps in the understanding of individual feature contributions, explaining the model's performance. Model generalization is ensured through 100-fold randomization and cross-validation during the performance evaluation.

The major contributions are as follows:

1. Preprocessing: employing morphological algorithms, filtering, and image sharpening during preprocessing effectively eliminates noise and artifacts from dermoscopic images. This process enhances image quality and accurately highlights skin lesion regions;
2. Comprehensive Feature Extraction: the method extracts shape features from processed images using techniques like GLMC for texture, color moments for color, and morphology for a holistic representation of lesion attributes;
3. Interpretable Feature Selection: the designed RFE feature selection method employs a 100-fold randomization strategy to derive the feature importance ranking. This ranking concurrently serves to elucidate the individual feature contributions. Consequently, this methodology inherently favors the selection of higher-ranked features, thereby ensuring a blend of enhanced credibility and interpretability;
4. Model Performance: through feature screening and model evaluation, including tenfold cross-validation and 100-fold randomization, the model's generalization is rigorously ensured.

2. Methodology

The CAD system melanoma classification method includes a preprocessing stage, feature selection stage, and classification stage. The preprocessing improved the quality of the dermoscopic image, and the skin lesions were segmented and cropped. The extracted texture, color, and shape features (TCS feature set) were selected, and internal ranking contributions were carefully analyzed and selected (RFE-TCS feature set). Finally, SVM ("polynomial(poly)", "radial basis kernel functions (rbf)" and "linear", "sigmoid"), logistic regression (LR), and Gaussian naive Bayes (GaussianNB) were used to test and train the selected optimization group. The detailed steps are shown in Figure 1.

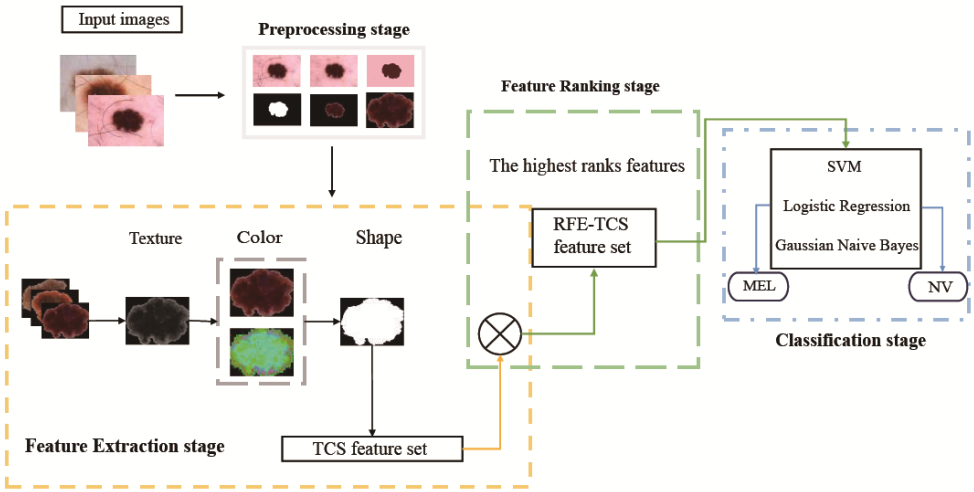


Figure 1. Block diagram of the proposed melanoma classification method.

2.1. Preprocessing of Dermoscopic Images

From the above, dermoscopic images are affected mainly by noise such as hairs and bubbles, edge, and contrast problems. For this purpose, morphological methods [20], filters as well as sharpening were used to solve these problems and improved the quality of the image. Considering the possible damage to the original image during the removal of hairs etc., a masking technique was used to perform some restoration process. Figure 2 shows the comparison of dermoscopic images before and after processing, and the above-mentioned problems such as noise were improved.

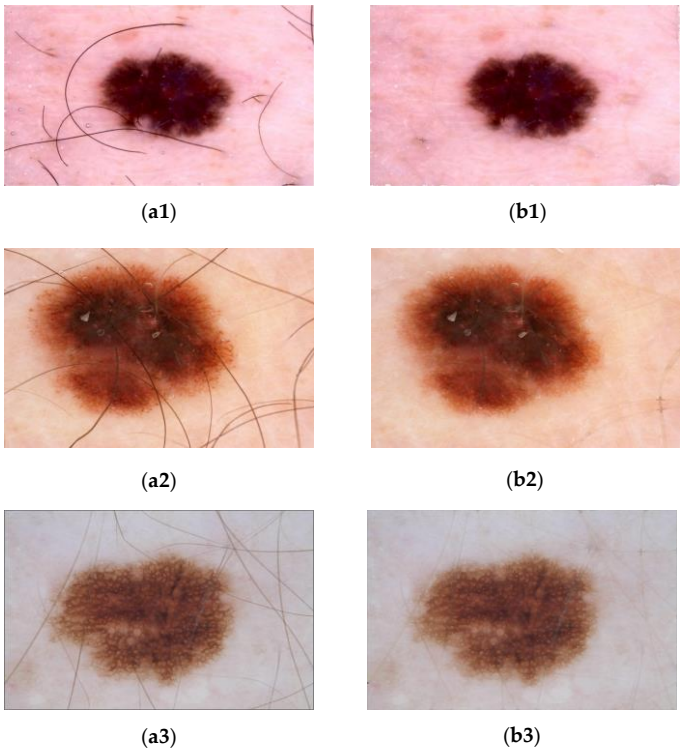


Figure 2. (**a1**–**a3**) The original images; (**b1**–**b3**) noise removed, sharpened and filtered images.

The k-means clustering algorithm is an unsupervised clustering technique based on partitioning, which is known for its fast convergence and easy implementation [21]. The k-means clustering algorithm is known to provide locally optimal solutions [22]. It is suitable for the segmentation of color images, including the classification of skin lesion areas and background regions in an image. In dermoscopic images, there is a clear color contrast between lesion areas and normal skin areas. Typically, diseased skin appears brown or black, while normal skin appears white or yellow. Given this difference, the RGB color space is effective in identifying skin lesion regions. Therefore, the k-means clustering algorithm was used for RGB color-based segmentation of skin fluoroscopy images [23].

The k-means clustering [24] algorithm was used to segment dermoscopic images with two initial cluster centers representing skin lesions and normal skin regions, respectively. The image masking technique was then used to extract the skin lesion region and ignore the background. To further minimize the background effect, the skin lesion region of the segmented dermoscopic image was cropped. Figure 3 illustrates the segmentation and cropping process.

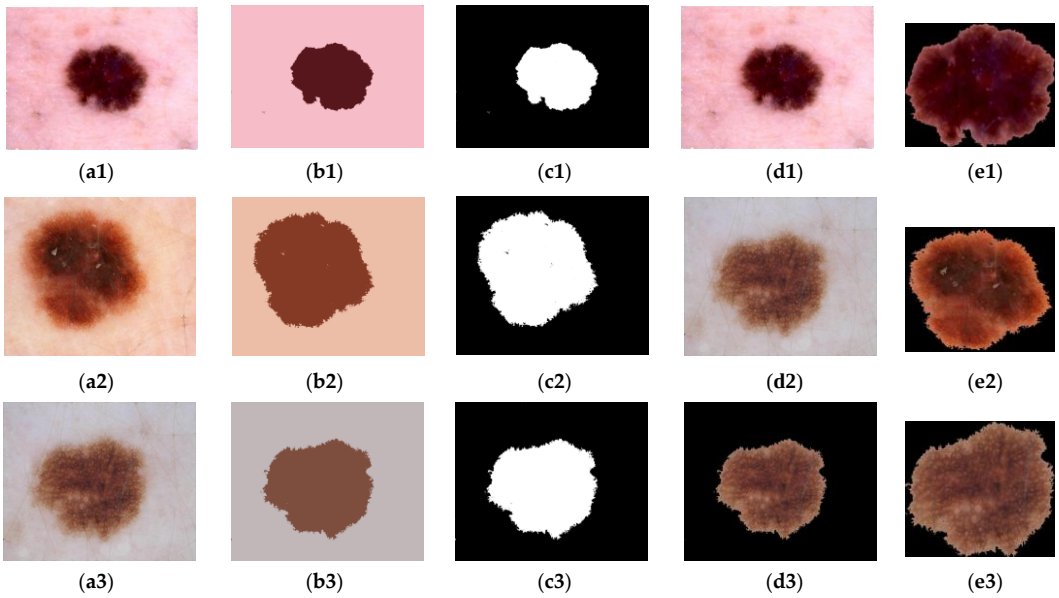


Figure 3. Results of the thresholding stage: (**a1**–**a3**) the original images; (**b1**–**b3**) the images after k-means; (**c1**–**c3**) the masks; (**d1**–**d3**) the lesions; (**e1**–**e3**) the cropped images.

2.2. Feature Extraction

2.2.1. Texture Feature

Texture characterization [25,26] captures changes in surface or structural patterns in an image. Gray scale covariance matrix (GLCM) [27,28] is a fundamental method to analyze texture features, which accurately reveals the roughness and repetition direction of the texture. In total, six feature parameters were selected: Contrast, Similarity, Homogeneity, Energy, Correlation, and Angular Second Matrix (ASM). The covariance matrix can have features extracted from scanned images with different orientation angles, as well as from images with different gray levels. The formulas [25–28] are shown in (1)–(6). To explore the effect of this, four direction angles (0°, 45°, 90°, and 135°) and three gray levels (8, 16, and 32) were selected.

Contrast quantifies the change in intensity within an image and represents the difference between the intensity of one pixel compared to another. Mathematically, it is calculated as follows:

$$\text{Contrast} = \sum_i \sum_j P(i,j) * |i-j|^2 \quad (1)$$

Dissimilarity assesses the level of distinction between pairs of pixels with varying gray levels in an image. It is calculated as follows:

$$\text{Dissimilarity} = \sum_i \sum_j P(i,j) * |i-j| \quad (2)$$

Angular second moment gauges the uniformity of the distribution of gray levels and the texture thickness within an image. This metric is calculated as follows:

$$\text{ASM} = \sum_i \sum_j P(i,\ j)^2 \quad (3)$$

Homogeneity quantifies the extent of variation in the textual components of the image, particularly in its uniformity. It is calculated using the following formula:

$$\text{Homogeneity} = \sum_i \sum_j \frac{P(i,\ j)}{1+|i-j|^2} \quad (4)$$

Energy measures the stability of gray level variations within the texture of an image. The calculation for energy is as follows:

$$\text{Energy} = \sum_i \sum_j P(i, j) \tag{5}$$

Correlation evaluates the similarity of gray levels within an image, either along rows or columns. The calculation for correlation is as follows:

$$\text{Correlation} = \sum_i \sum_j \frac{(i - \text{Mean}) * (j - \text{Mean}) * P(i,j)^2}{\text{Variance}} \tag{6}$$

In calculating the texture features, the variables "i" and "j" represent the coordinates of the pixels in the image are the i-axis and j-axis in the GLCM, and P(i,j) denotes the frequency of occurrence at pixels at a fixed location. The "mean" refers to the average value in the texture feature, while the "variance" represents the dispersion or spread of the values in the texture feature.

2.2.2. Color Feature

Color moments were used in [29] to extract the color features of the skin lesion region in dermoscopic images. The parameters are first moments, second moments, and third moments of color. The followings are the Formulas (7)–(9) for the calculation [30] of color moments:

The first moment quantifies the sensitivity of an image, reflecting its overall intensity distribution. It is calculated as follows:

$$\text{First moment} = \sum_j P(i, j) \tag{7}$$

The second moment provides insight into the range of the color distribution within an image and offers information about patterns and contrasts. The second moment is calculated as follows:

$$\text{Second moment} = \left(\frac{1}{N} \sum_j \left(P(i, j) - u_i \right)^2 \right)^{\frac{1}{2}} \tag{8}$$

The third moment conveys the symmetry of color distribution in an image, indicating the balance and arrangement of colors. It is calculated as follows:

$$\text{Third moment} = \left(\frac{1}{N} \sum_j \left(P(i, j) - u_i \right)^3 \right)^{\frac{1}{3}} \tag{9}$$

In the image, the variables "i" and "j" represent the coordinates of the pixels while "N" represents the total number of sub-pixels in the image. The six color channels R, G, B, H, S, V were separated under the RGB color space and HSV color space, respectively, and then color features in each single-color channel were extracted using color moments.

2.2.3. Shape Feature

The shape feature is a visual description of the lesion area and includes various metrics such as area and perimeter [30]. The area feature (A) quantifies the total number of pixels within the lesion area, while the perimeter feature (P) quantifies the total number of contour pixels on the boundary of the lesion area. Based on these two parameters, other shape descriptors, such as dispersion, saturation, and roundness, can be derived, as shown in (10)–(12) [31]. In the shape characterization process, dispersion, saturation, and roundness were chosen as descriptors to fully present the shape characteristics of the lesion area.

Dispersion characterizes the extent of spread within a region in an image. It provides information about how the elements within the region are distributed. Dispersion is the

ratio of the square of the perimeter of the lesion area to the area and describes the process of regional dispersion. The calculation for dispersion is as follows:

$$\text{Dispersity} = \frac{P^2}{A} \tag{10}$$

Saturation, which can also be referred to as convexity, pertains to the shape of the region's boundary in an image. It describes how closely the region resembles a convex shape. Saturation is the ratio of area to perimeter of the lesion region. The calculation for saturation (or convexity) is as follows:

$$\text{Saturation} = \frac{A}{P} \tag{11}$$

Circularity evaluates how closely the shape of a region resembles a circle, indicating the level of proximity to a circular form. It provides insights into the compactness of the region. Roundness describes the similarity of the shape of the lesion region to a circle. The calculation for circularity is as follows:

$$\text{Roundness} = \frac{(4 * \Pi * A)}{P^2} \tag{12}$$

2.3. Feature Selection by RFE Ranking

According to the above, a total of 93 features (72 texture features, 18 color features, and 3 shape features) were extracted from the skin lesion area. These features were combined into a feature vector matrix named TCS for melanoma classification. To simplify the computational process, the range of values for the color and shape parameters was narrowed down.

All these features were combined into a feature vector matrix named TCS for melanoma classification. RFE was used to select features and determine their importance. RFE selected the best set of features by eliminating and shifting. The best features were wrapped after repetition [4]. On the other hand, the importance and contribution of each feature were obtained for each elimination [31]. For this reason, the features could be ranked according to the size of the contribution. To select features more reliably, a 100-fold randomization method was designed. The dataset was shuffled before each training. The final ranking was calculated based on the entire ranking. The features with high contribution were selected and combined into a new feature vector matrix, called RFE-TCS, which was specifically used for melanoma classification.

2.4. Methodological Process Design

SVM [32] is a kind of supervised machine learning method widely used for classification and classification tasks [33]. It is particularly effective for binary classification problems. The goal is to identify support vectors, i.e., subsets of samples that are farthest from the hyperplane and represent different classes.

One of the strengths of SVM [34] is its stability, even with small sample sizes. It can handle different classification problems and optimize the classification results by choosing appropriate kernel functions. For this experiment, "poly", "rbf", "linear" and "sigmoid" were chosen to train and classify the experimental dataset to obtain the best classification model.

Logistic regression [35] serves as a widely employed classification algorithm that is particularly suitable for binary classification tasks. Hence, it is embraced as a classifier model. Additionally, naive Bayes classifiers employ Bayes' theorem for classification. Thus, the Gaussian naive Bayes classifier was selected [36].

The entire experiment was conducted using Python, utilizing machine learning techniques. The training process consists of using the extracted features to train the melanoma classification model. The workflow of the machine learning process is shown in Figure 4.

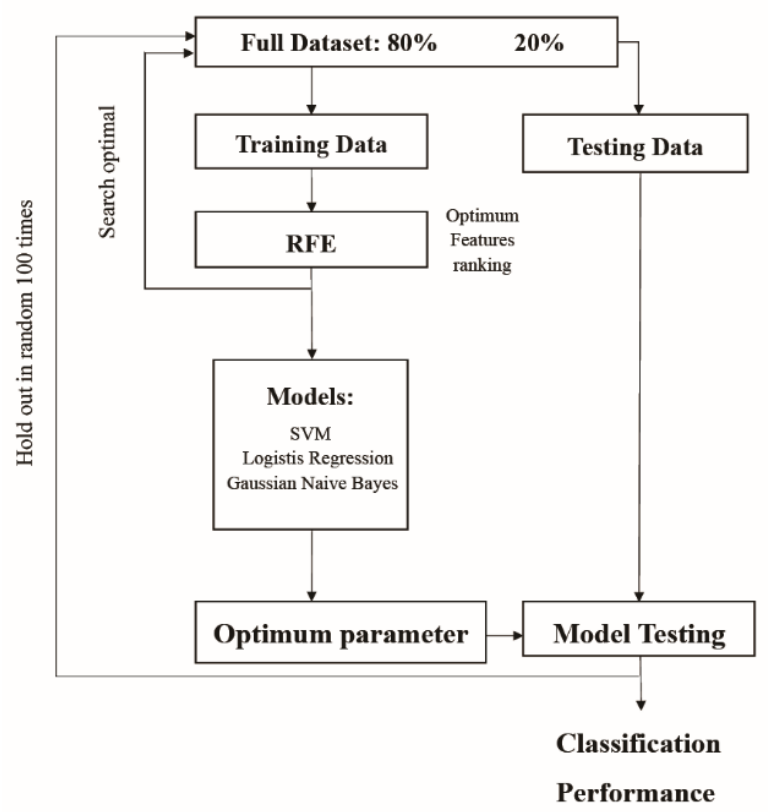


Figure 4. The flowchart of machine learning. RFE was used to rank the importance of features. The classification algorithms use SVM (poly, rbf, linear, sigmoid), LR and GaussianNB. This method was randomly repeated 100 times, with 80% of the data used for training and 20% for testing.

The entire experimental procedure was executed using Python, employing various machine learning techniques. The training phase involved utilizing the extracted features to train the melanoma classification model. The workflow of this machine learning process is visually depicted in Figure 4. The dataset was divided into two subsets: an 80% portion earmarked as the training set, and a remaining 20% portion designated as the test set:

(1) The initial step encompassed the utilization of all feature sets (TCS) as input features for the machine learning process. The recursive feature elimination (RFE) algorithm was employed for feature ranking and selection. To facilitate an internal visualization of these features, a 100-fold randomization method was employed for carrying out the statistical analysis. The outcomes of the feature selection process were then subjected to comprehensive statistical analysis, culminating in the derivation of a ranking indicating the importance of each feature. Based on this ranking, features with high importance were selectively chosen to create a novel feature set termed "RFE-TCS";
(2) The repertoire of machine learning methods encompassed SVM (using the polynomial kernel, radial basis function kernel, linear kernel, and sigmoid kernel), logistic regression, and a Gaussian Bayesian classifier;
(3) This divisional process was repeated 100 times, with each instance involving the random shuffling of the dataset. The models were all cross-validated using tenfold cross-validation. As part of the evaluation protocol, performance metrics such as average classification accuracy and the area under the curve (AUC) were meticulously computed.

3. Results and Discussion

In the feature selection process, each random shuffling would produce a round of optimized combinations, statistics of the optimized combinations of 100 times the results. The number of times a feature appeared in an optimized combination was recorded as the

frequency, and the ranking of features was counted and computationally analyzed based on all the results.

In the classification training stage, the evaluation metrics used to assess the performance of the CAD system-based melanoma lesion identification method were accuracy, sensitivity, and specificity. Also, the ROC-AUC [37] curve was drawn. These metrics are commonly used in medical image analysis to measure the effectiveness of classification models. The formulas for accuracy, sensitivity, and specificity [38] are as shown in (13)–(15):

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{13}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{14}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{15}$$

where 'TP' is the number of correctly identified melanoma lesions as melanoma, 'TN' is the number of correctly identified benign lesions as non-melanoma, 'FP' is the number of benign lesions incorrectly identified as melanoma, and 'FN' is the number of melanoma lesions incorrectly identified as benign lesions.

The ISIC dataset is widely recognized as the largest public database for research in dermoscopic image analysis. In the experiment, a total of 200 dermoscopic images of melanoma and 200 dermoscopic images of benign nevus were selected from the ISIC 2019 database. This dataset consists of 400 images in total, with an equal representation of both melanoma and benign nevus cases. The balanced dataset allows for the fair and reliable evaluation of the CAD system's performance in distinguishing between melanoma and benign lesions.

The PH^2 dataset, introduced and detailed by Mendonca et al., comprises 200 dermoscopic images. This dataset categorically segregates images into melanomas and benign nevi.

Melanoma has a relatively similar presentation with benign nevi [39], and the classification between melanoma and benign nevi is mainly designed to minimize misclassification.

3.1. Comparison of Melanoma Classification Performance

Table 1 displays the performance evaluation of classifiers, utilizing the original feature set (TCS) as the training input. The average classification accuracy (using the 100-fold randomization method) ranked as follows: LR (ACC = 74.71%, AUC = 0.767) = SVM linear (ACC = 75.70%, AUC = 0.756) > GaussianNB (ACC = 68.96%, AUC = 0.703) > SVM poly (ACC = 65.92%, AUC = 0.664) > SVM sigmoid (ACC = 62.93%, AUC = 0.566) > SVM rbf (ACC = 61.96%, AUC = 0.67). The findings underscore the LR classifier's supremacy over its counterparts, emerging as the frontrunner in terms of performance. Boasting an accuracy of 74.71% (ACC) and an AUC of 0.767, the logistic regression classifier excels in its capacity to discriminate between classes. Furthermore, this classifier shows enhancements in both sensitivity and specificity.

Table 1. Performance of original features in different classifiers.

Classifier	Feature Set	ACC (Average) (%)	Sensitivity	Specificity	AUC
SVM poly	TCS	65.92	0.568	0.760	0.664
SVM rbf	TCS	61.96	0.732	0.608	0.670
SVM linear	TCS	74.71	0.72	0.8	0.756
SVM sigmoid	TCS	62.93	0.664	0.567	0.566
Logistic Regression	TCS	74.71	0.735	0.799	0.767
GaussianNB	TCS	68.96	0.616	0.788	0.703

Table 2 displays the performance appraisal shifts its focus to the classifier's efficacy when leveraging the selected feature set (RFE-TCS) as the training input. The average classification accuracy (using the 100-fold randomization method) ranked as SVM rbf (ACC = 77.27%, AUC = 0.766) > SVM linear (ACC = 74.88%, AUC = 0.748) > LR (ACC = 74.80%, AUC = 0.749) > SVM poly (ACC = 74.73%, AUC = 0.766) > SVM sigmoid (ACC = 73.68%, AUC = 0.745) > GaussianNB (ACC = 69.13%, AUC = 0.691), the SVM classifier with the rbf kernel function distinguished itself by surpassing alternative classifiers, thereby attaining the best performance with an accuracy of 76.47% and AUC of 0.766.

Table 2. Performance of selected features in different classifiers.

Classifier	Feature Set	ACC (Average) (%)	Sensitivity	Specificity	AUC
SVM poly	RFE-TCS	74.73	0.69	0.85	0.766
SVM rbf	RFE-TCS	77.27	0.69	0.85	0.766
SVM linear	RFE-TCS	74.88	0.656	0.84	0.748
SVM sigmoid	RFE-TCS	73.68	0.656	0.83	0.745
Logistic Regression	RFE-TCS	74.80	0.694	0.804	0.749
GaussianNB	RFE-TCS	69.13	0.609	0.774	0.691

Table 3 shows the classification results on PH2 using the filtered feature set (RFE-TCS). From the table, it can be seen that the SVM classifier with the linear kernel function achieved an accuracy of 82.42%.

Table 3. Performance of selected features in PH2 dataset.

Classifier	ACC (Average) (%)
SVM poly	81.35
SVM rbf	79.60
SVM linear	81.92
SVM sigmoid	80.41
Logistic Regression	80.77
GaussianNB	73.12

3.2. RFE Ranking Explains the Importance of Features

The results of ranking the importance of texture, color and shape features in melanoma identification using RFE method are shown in Figure 5. This ranking is the feature importance ranking. From the ranking in Figure 5, it is shown that texture features occupy the major high ranking. In Popecki's study [40], although the key role of texture features was observed, it was not specifically discussed or explained. The high or low feature importance ranking obtained by the designed methodology explains the reason for the texture feature as a key feature. This feature importance ranking method provides more insights and bridges the gap in feature selection in previous studies. Moreover, the ranking, obtained from the 100-fold randomization experiments that effectively highlighted the contribution of features, is also more convincing.

The performance results in Tables 1 and 2, Figures 6 and 7 affirm the filtered feature set's effectiveness, underscoring the optimized features' role in enhancing melanoma recognition. The RFE feature importance rankings are plausible and offer explanatory potential for model outcomes.

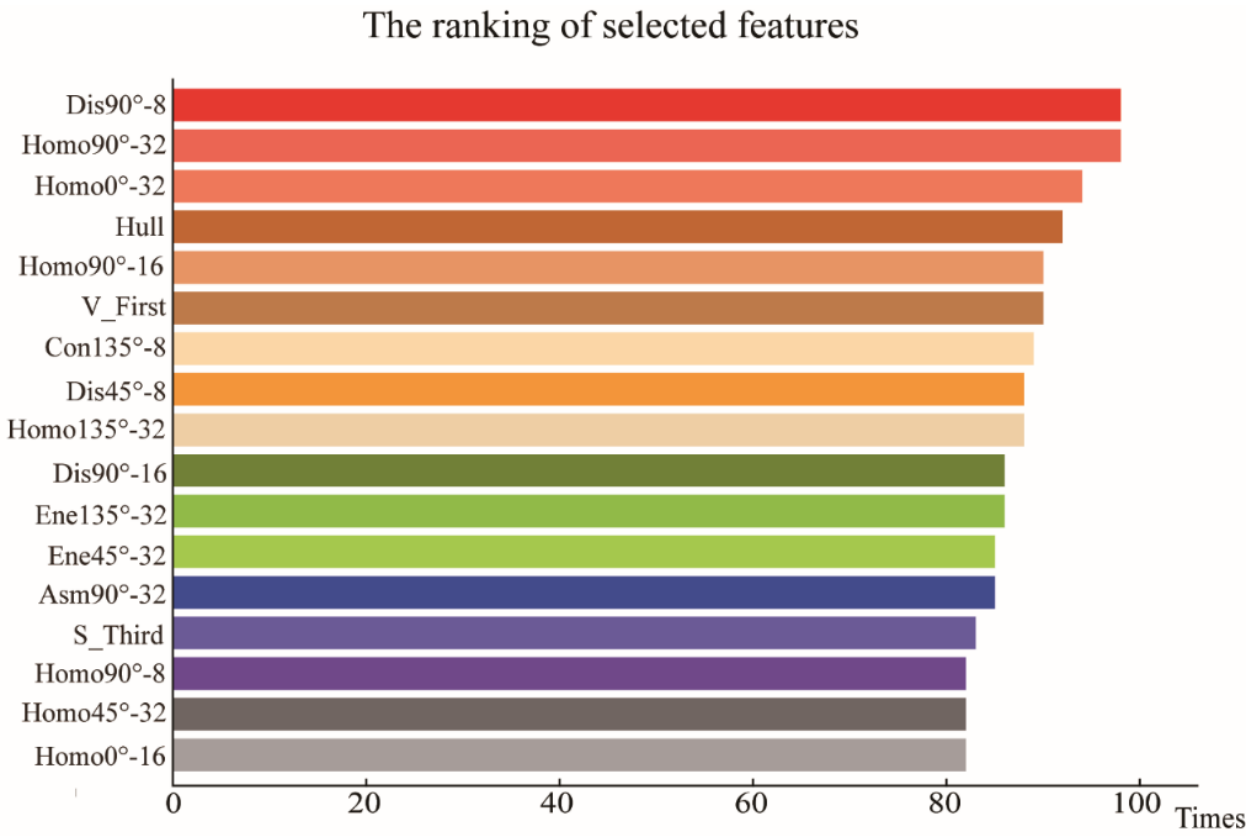


Figure 5. Ranking of features (top 17 features) from top to bottom: Dissimilarity90°-8, Homogeneit90°-32, Homogeneit0°-32, Hull, Homogeneit90°-16, V_First moment, Contrast135°-8, Dissimilarity45°-8, Homogeneit135°-32, Dissimilarity90°-16, Energy135°-32, Energy45°-32, Asm90°-32, S_ Third moment, Homogeneity90°-8, Homogeneit45°-32, Homogeneity0°-16. These rankings provide valuable information about the importance and contribution of each feature in the classification of melanoma. These were ranked by the 100-fold randomization method.

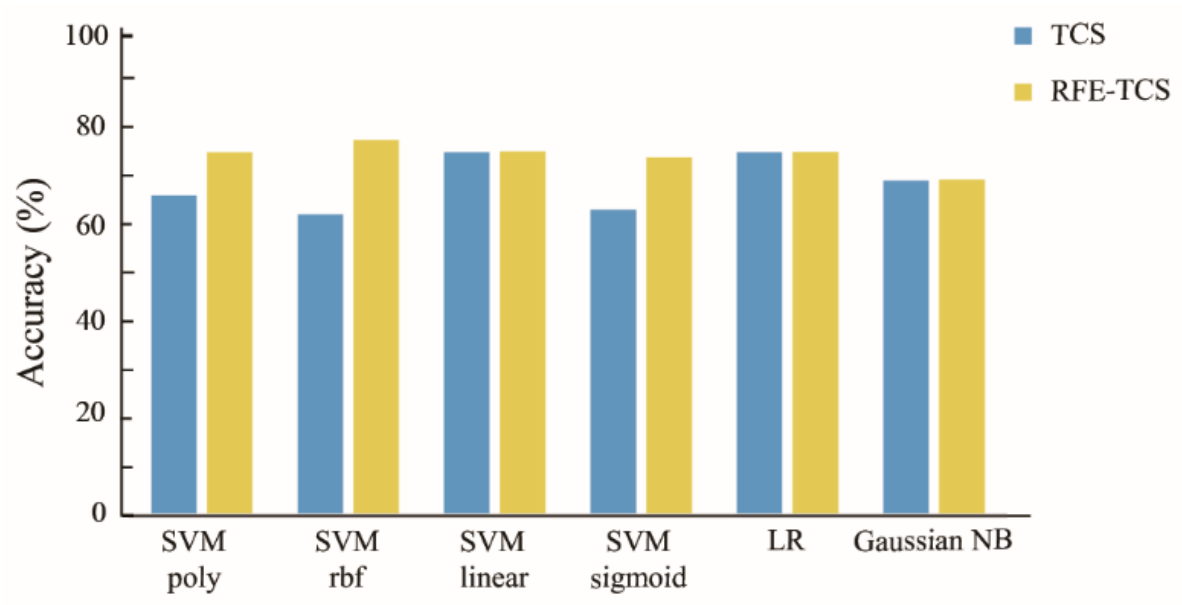


Figure 6. Comparison of classification performance of different algorithms by adjusting the RFE-TCS feature set.

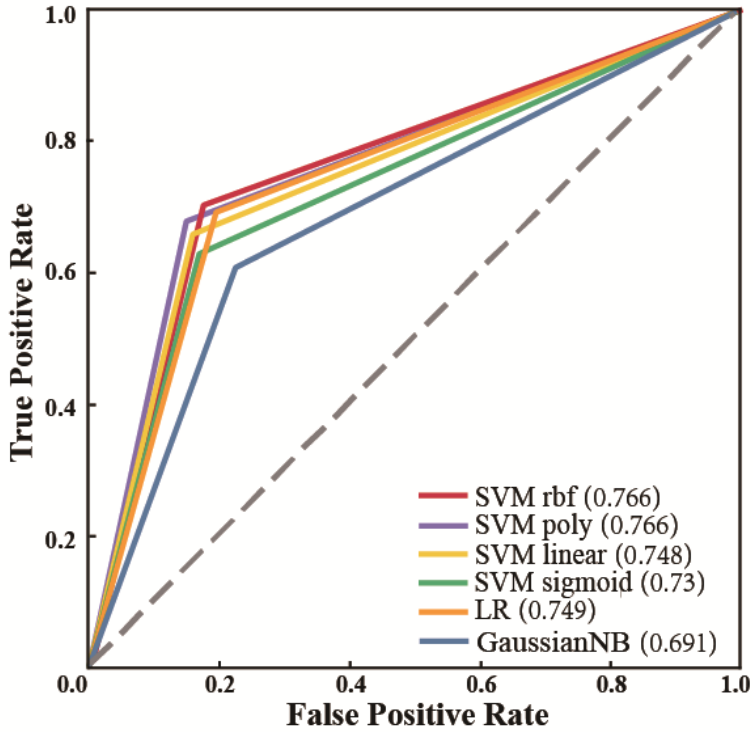


Figure 7. Classification performance and AUC for the randomized 100-fold randomization method.

3.3. Generalization Discussion of the Classifiers Models

In Table 3, the PH2 dataset was utilized for the experimental procedures. Specifically, the individual classifiers trained on the ISIC-selected dataset were employed to test the PH2 dataset. The results, as presented in the table, demonstrate consistent and stable performance across all individual classifiers. Particularly, each classifier exhibited an observable enhancement in accuracy. These findings serve as a positive indication of the generalized nature of the proposed method's model.

To evaluate the generalization capability of SVM models, the application of cross-validation techniques is a common practice [41,42]. Prior studies have effectively demonstrated the generalization achieved through cross-validation [43] by training SVM models on a discovery dataset and subsequently assessing their performance using a distinct replicated dataset [44]. In the approach, internal tenfold cross-validation [45] was employed to optimize model performance.

Furthermore, the model's capacity to generalize was further assessed using an external method involving 100 times randomization replicate instances, which utilized a holdout approach. This methodology was chosen to enhance control over the model's generalization capability and provide a comprehensive evaluation of its overall performance.

3.4. Comparative Performance with Other Models

Table 4 presents a comparative assessment of the accuracy achieved by other different models in the context of melanoma identification. In the existing literature, GaussianNB [33], LR [46], RF [47], and KNN [48] classifiers have been reported to achieve accuracy rates of 65.93%, 72%, 74.28%, and 75.00% respectively. The proposed SVM model, employing the specially designed radial basis function (rbf) kernel, exhibits higher recognition accuracy when compared to other prevalent machine learning classifiers.

In addition to its capacity for automated melanoma recognition, the proposed method goes a step further by visually illustrating the contributions of individual features. This visualization not only bolsters confidence in the model but also incorporates the use of randomized experiments to ensure the model's generalizability. This distinctive approach sets it apart from the other methods, providing a transparent and reliable model. Some

studies [49] have explored the use of deep learning neural networks in achieving greater accuracy. This is an approach that may be considered in the future.

Table 4. Performance comparisons of proposed model with other recent models.

Model	Accuracy (%)
GaussianNB [33]	65.93
LR [46]	72.00
RF [47]	74.28
KNN [48]	75.00
Proposed GaussianNB	69.13
Proposed LR	74.80
Proposed SVM	77.27

3.5. Limitations

Although features were extended in a number of ways, there are still some other approaches to feature selection that can be adopted in future work. In terms of model performance, comparing the previous studies LR and GaussianNB have a relatively better surface. However, although the performance of the model is better than SVM [48], the accuracy can still be improved. There is still room for improvement in terms of the final model's performance; the next step will be the use of different modelling design methods to improve the accuracy of results.

4. Conclusions

This study presents an interpretable machine learning melanoma recognition method that emphasizes the importance of features. The method starts with pre-processing, feature extraction (texture features using GLCM, color features using color moments, and shape features using morphological techniques) and feature ranking of dermoscopic images. Subsequently, SVM classifiers (poly, rbf, linear, sigmoid), logistic regression and Gaussian Bayesian machine learning techniques are used for training.

An important aspect of the method is the visual ranking of feature contributions by RFE during the feature-extraction phase. The obtained feature importance ranking explains the importance of texture features. In addition, the feature selection is more convincing. Another aspect is that in the classification stage, the model is cross-validated ten times and 100 randomized experiments are used to enhance the credibility of the results and the generalization ability of the model.

The results confirm that the feature set optimized by feature importance ranking performs better. The model demonstrated significant generalization ability on different datasets. This approach provides insights into melanoma identification and emphasizes the value of feature selection and interpretation for robust model design.

Author Contributions: Conceptualization, Methodology, Writing—review and editing, Z.L.; Software, Formal analysis, Writing—original draft, Q.J.; Methodology, Formal analysis, X.Y. and Y.Z.; Data curation, S.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the key specialized research and development breakthrough of Henan province (Grant No. 232102210030 to Y.Z.).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The ISIC dataset utilized in this work is openly available at: https://challenge.isic-archive.com/data. The PH2 dataset utilized in this work is openly available at: https://www.fc.up.pt/addi/ph2%20database.html.

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

References

1. Narayanan, D.L.; Saladi, R.N.; Fox, J.L. Ultraviolet radiation and skin cancer. *Int. J. Dermatol.* **2010**, *49*, 978–986. [CrossRef] [PubMed]
2. Skin Cancers. Available online: http://www.who.int/uv/faq/skincancer/en/index1.html (accessed on 15 January 2020).
3. Shchetinin, E.Y.; Demidova, A.V.; Kulyabov, D.S.; Sevastyanov, L.A. Skin Lesion Classification Using Deep Learning Methods. *Math. Biol. Bioinform.* **2020**, *15*, 180–194. [CrossRef]
4. Chatterjee, S.; Dey, D.; Munshi, S. Integration of morphological preprocessing and fractal-based feature extraction with recursive feature elimination for skin lesion type classification. *Comput. Methods Programs Biomed.* **2019**, *178*, 201–218. [CrossRef] [PubMed]
5. Adegun, A.A.; Viriri, S. FCN-Based DenseNet Framework for Automated Detection and Classification of Skin Lesions in Dermoscopy Images. *IEEE Access* **2020**, *8*, 150377–150396. [CrossRef]
6. Giotis, I.; Molders, N.; Land, S.; Biehl, M.; Jonkman, M.F.; Petkov, N. MED-NODE: A computer-assisted melanoma diagnosis system using non-dermoscopic images. *Expert. Syst. Appl.* **2015**, *42*, 6578–6585. [CrossRef]
7. Joseph, S.; Olugbara, O.O. Preprocessing Effects on Performance of Skin Lesion Saliency Segmentation. *Diagnostics* **2022**, *12*, 344. [CrossRef]
8. Ashraf, H.; Waris, A.; Ghafoor, M.F.; Gilani, S.O.; Niazi, I.K. Melanoma segmentation using deep learning with test-time augmentations and conditional random fields. *Sci. Rep.* **2022**, *12*, 3948. [CrossRef] [PubMed]
9. Bakheet, S.; Al-Hamadi, A. Computer-aided diagnosis of malignant melanoma using Gabor-based entropic features and multilevel neural networks. *Diagnostics* **2020**, *10*, 822. [CrossRef]
10. Garnavi, R.; Aldeen, M.; Celebi, M.E.; Varigos, G.; Finch, S. Border detection in dermoscopy images using hybrid thresholding on optimized color channels. *Comput. Med. Imaging Graph.* **2011**, *35*, 105–115. [CrossRef]
11. Yu, L.; Chen, H.; Dou, Q.; Qin, J.; Heng, P.A. Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks. *IEEE Trans. Med. Imaging* **2017**, *36*, 994–1004. [CrossRef]
12. Hugo, W.; Zaretsky, J.M.; Sun, L.; Song, C.; Moreno, B.H.; Hu-Lieskovan, S.; Berent-Maoz, B.; Pang, J.; Chmielowski, B.; Cherry, G.; et al. Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. *Cell* **2016**, *165*, 35–44. [CrossRef] [PubMed]
13. Rastgoo, M.; Garcia, R.; Morel, O.; Marzani, F. Automatic differentiation of melanoma from dysplastic nevi. *Comput. Med. Imag. Graph.* **2015**, *43*, 44–52. [CrossRef] [PubMed]
14. Bharathi, G.; Malleswaran, M.; Muthupriya, V. Detection and diagnosis of melanoma skin cancers in dermoscopic images using pipelined internal module architecture (PIMA) method. *Microsc. Res. Tech.* **2023**, *86*, 701–713. [CrossRef] [PubMed]
15. Nasir, M.; Attique Khan, M.; Sharif, M.; Lali, I.U.; Saba, T.; Iqbal, T. An improved strategy for skin lesion detection and classification using uniform segmentation and feature selection based approach. *Microsc. Res. Tech.* **2018**, *81*, 528–543. [CrossRef] [PubMed]
16. Wahba, M.A.; Ashour, A.S.; Guo, Y.; Napoleon, S.A.; Elnaby, M.M.A. A novel cumulative level difference mean based GLDM and modified ABCD features ranked using eigenvector centrality approach for four skin lesion types classification. *Comput. Methods Programs Biomed.* **2018**, *165*, 163–174. [CrossRef]
17. Moldovanu, S.; Damian Michis, F.A.; Biswas, K.C.; Culea-Florescu, A.; Moraru, L. Skin Lesion Classification Based on Surface Fractal Dimensions and Statistical Color Cluster Features Using an Ensemble of Machine Learning Techniques. *Cancers* **2021**, *13*, 5256. [CrossRef] [PubMed]
18. Lin, X.; Li, C.; Zhang, Y.; Su, B.; Fan, M.; Wei, H. Selecting Feature Subsets Based on SVM-RFE and the Overlapping Ratio with Applications in Bioinformatics. *Molecules* **2018**, *23*, 52. [CrossRef]
19. Sanz, H.; Valim, C.; Vegas, E.; Oller, J.M.; Reverter, F. SVM-RFE: Selection and visualization of the most relevant features through non-linear kernels. *BMC Bioinform.* **2018**, *19*, 432. [CrossRef]
20. Rehman, A.; Khan, M.A.; Mehmood, Z.; Saba, T.; Sardaraz, M.; Rashid, M. Microscopic melanoma detection and classification: A framework of pixel-based fusion and multilevel features reduction. *Microsc. Res. Tech.* **2020**, *83*, 410–423. [CrossRef]
21. Vaiyapuri, T.; Balaji, P.S.S.; Alaskar, H.; Sbai, Z. Computational Intelligence-Based Melanoma Detection and Classification Using Dermoscopic Images. *Comput. Intell. Neurosci.* **2022**, *2022*, 2370190. [CrossRef]
22. Nawaz, M.; Mehmood, Z.; Nazir, T.; Naqvi, R.A.; Rehman, A.; Iqbal, M.; Saba, T. Skin cancer detection from dermoscopic images using deep learning and fuzzy k-means clustering. *Microsc. Res. Tech.* **2022**, *85*, 339–351. [CrossRef] [PubMed]
23. Dash, M.; Londhe, N.D.; Ghosh, S.; Shrivastava, V.K.; Sonawane, R.S. Swarm intelligence-based clustering technique for automated lesion detection and diagnosis of psoriasis. *Comput. Biol. Chem.* **2020**, *86*, 107247. [CrossRef]
24. Chang, C.C.; Li, Y.Z.; Wu, H.C.; Tseng, M.H. Melanoma Detection Using XGB Classifier Combined with Feature Extraction and K-Means SMOTE Techniques. *Diagnostics* **2022**, *12*, 1747. [CrossRef] [PubMed]
25. Zareen, S.S.; Guangmin, S.; Li, Y.; Kundi, M.; Qadri, S.; Qadri, S.F.; Ahmad, M.; Khan, A.H. A Machine Vision Approach for Classification of Skin Cancer Using Hybrid Texture Features. *Comput. Intell. Neurosci.* **2022**, *2022*, 4942637. [CrossRef]
26. Rastghalam, R.; Danyali, H.; Helfroush, M.S.; Celebi, M.E.; Mokhtari, M. Skin Melanoma Detection in Microscopic Images Using HMM-Based Asymmetric Analysis and Expectation Maximization. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 3486–3497. [CrossRef] [PubMed]
27. Talavera-Martínez, L.; Bibiloni, P.; González-Hidalgo, M. Computational texture features of dermoscopic images and their link to the descriptive terminology: A survey. *Comput. Methods Programs Biomed.* **2019**, *182*, 105049. [CrossRef]

28. Wahba, M.A.; Ashour, A.S.; Napoleon, S.A.; Abd Elnaby, M.M.; Guo, Y. Combined empirical mode decomposition and texture features for skin lesion classification using quadratic support vector machine. *Health Inf. Sci. Syst.* **2017**, *5*, 10. [CrossRef]
29. Mishra, N.K.; Kaur, R.; Kasmi, R.; Hagerty, J.R.; LeAnder, R.; Stanley, R.J.; Moss, R.H.; Stoecker, W.V. Automatic lesion border selection in dermoscopy images using morphology and color features. *Ski. Res. Technol.* **2019**, *25*, 544–552. [CrossRef]
30. He, W.; Liu, T.; Han, Y.; Ming, W.; Du, J.; Liu, Y.; Yang, Y.; Wang, L.; Jiang, Z.; Wang, Y.; et al. A review: The detection of cancer cells in histopathology based on machine vision. *Comput. Biol. Med.* **2022**, *146*, 105636. [CrossRef]
31. Pitchiah, M.S.; Rajamanickam, T. Efficient feature-based melanoma skin image classification using machine learning approaches. *Trait. Signal* **2022**, *39*, 1663–1671. [CrossRef]
32. Zhang, J.; Ding, Q.; Li, X.L.; Hao, Y.W.; Yang, Y. Support Vector Machine versus Multiple Logistic Regression for Prediction of Postherpetic Neuralgia in Outpatients with Herpes Zoster. *Pain Physician* **2022**, *25*, E481–E488.
33. Seeja, R.D.; Suresh, A. Deep Learning Based Skin Lesion Segmentation and Classification of Melanoma Using Support Vector Machine (SVM). *Asian Pac. J. Cancer Prev.* **2019**, *20*, 1555–1561.
34. Varalakshmi, P.; Devi, V.A.; Ezhilarasi, M.; Sandhiya, N. Enhanced Dermatoscopic Skin Lesion Classification using Machine Learning Techniques. In Proceedings of the 2021 Sixth International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, India, 25–27 March 2021; pp. 68–71.
35. Yang, T.Y.; Chien, T.W.; Lai, F.J. Web-Based Skin Cancer Assessment and Classification Using Machine Learning and Mobile Computerized Adaptive Testing in a Rasch Model: Development Study. *JMIR Med. Inform.* **2022**, *10*, e33006. [CrossRef] [PubMed]
36. Shetty, B.; Fernandes, R.; Rodrigues, A.P.; Chengoden, R.; Bhattacharya, S.; Lakshmanna, K. Skin lesion classification of dermoscopic images using machine learning and convolutional neural network. *Sci. Rep.* **2022**, *12*, 18134. [CrossRef] [PubMed]
37. Bassel, A.; Abdulkareem, A.B.; Alyasseri, Z.A.A.; Sani, N.S.; Mohammed, H.J. Automatic Malignant and Benign Skin Cancer Classification Using a Hybrid Deep Learning Approach. *Diagnostics* **2022**, *12*, 2472. [CrossRef]
38. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **2017**, *542*, 115–118, Erratum in: *Nature* **2017**, *546*, 686. [CrossRef]
39. Casadonte, R.; Kriegsmann, M.; Kriegsmann, K.; Hauk, I.; Meliß, R.R.; Müller, C.S.L.; Kriegsmann, J. Imaging Mass Spectrometry-Based Proteomic Analysis to Differentiate Melanocytic Nevi and Malignant Melanoma. *Cancers* **2021**, *13*, 3197. [CrossRef]
40. Popecki, P.; Jurczyszyn, K.; Ziętek, M.; Kozakiewicz, M. Texture Analysis in Diagnosing Skin Pigmented Lesions in Normal and Polarized Light-A Preliminary Report. *J. Clin. Med.* **2022**, *11*, 2505. [CrossRef]
41. Shakeel, C.S.; Khan, S.J.; Chaudhry, B.; Aijaz, S.F.; Hassan, U. Classification Framework for Healthy Hairs and Alopecia Areata: A Machine Learning (ML) Approach. *Comput. Math. Methods Med.* **2021**, *2021*, 1102083. [CrossRef]
42. Zafar, K.; Gilani, S.O.; Waris, A.; Ahmed, A.; Jamil, M.; Khan, M.N.; Sohail Kashif, A. Skin Lesion Segmentation from Dermoscopic Images Using Convolutional Neural Network. *Sensors* **2020**, *20*, 1601. [CrossRef]
43. Vakharia, V.; Shah, M.; Nair, P.; Borade, H.; Sahlot, P.; Wankhede, V. Estimation of Lithium-ion Battery Discharge Capacity by Integrating Optimized Explainable-AI and Stacked LSTM Model. *Batteries* **2023**, *9*, 125. [CrossRef]
44. Kumar, S.M.; Kumanan, T. Skin lesion classification system using shearlets. *Comput. Syst. Sci. Eng.* **2023**, *44*, 833–844. [CrossRef]
45. Pohjankukka, J.; Pahikkala, T.; Nevalainen, P.; Heikkonen, J. Estimating the prediction performance of spatial models via spatial k-fold cross validation. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 2001–2019. [CrossRef]
46. Bechelli, S.; Delhommelle, J. Machine Learning and Deep Learning Algorithms for Skin Cancer Classification from Dermoscopic Images. *Bioengineering* **2022**, *9*, 97. [CrossRef] [PubMed]
47. Murugan, A.; Nair, S.A.H.; Preethi, A.A.P.; Kumar, K.P.S. Diagnosis of skin cancer using machine learning techniques. *Microprocess. Microsyst.* **2021**, *81*, 103727. [CrossRef]
48. Bakheet, S.; Alsubai, S.; El-Nagar, A.; Alqahtani, A. A Multi-Feature Fusion Framework for Automatic Skin Cancer Diagnostics. *Diagnostics* **2023**, *13*, 1474. [CrossRef]
49. Jasil, S.P.G.; Ulagamuthalvi, V. Deep learning architecture using transfer learning for classification of skin lesions. *J. Ambient Intell. Humaniz. Comput.* **2021**, 1–8. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.