Diagnosis of Melanoma and Nevus Using integrated Data for Argentina Population

Best of Both World

April 2024

Abstract

Melanoma is one of the most deadly types of skin cancer and the most effective treatment is to detect and remove the lesion at the earliest possible stage. In clinical practice, however, it is sometimes very difficult to distinguish melanoma from nevus because of the similarity of its symptoms and characteristics. Therefore, an accurate automatic diagnostic system is necessary. Current computer-aided skin lesion classification methods only focus on the image data of the lesion area and ignore the important role that the patient's personal information may play in the diagnosis process. In this paper, we propose a computer-aided melanoma diagnosis method that can use both image data and tabular data containing patients' personal information. The lesion area is first segmented using U-Net, and colour and texture features are extracted from the segmented image and the original image. These features are combined with the patient's personal information to form a super feature vector. XGBoost model is utilized for the classification of melanoma and nevus based on this feature vector. In this work, we use a skin lesion dataset collected in Argentina, which is the first dataset collected in Argentina for the evaluation of AI tools for skin lesion classification. Our method achieves a competitive 89% classification accuracy for melanoma and nevus on this dataset.

Keywords— Melanoma, nevus, skin lesion diagnosis, U-Net, XGBoost

1 Introduction

Melanoma is a highly deadly form of skin cancer[1]. If detected in time, removal of the cancerous area can completely cure a melanoma patient with a very good prognosis. However, if detected late, the cancer may have already spread throughout the body and can be life-threatening. Therefore, early detection of melanoma is very important for treatment. If melanoma is diagnosed at an early stage, the mortality rate can be reduced by up to 90%[2][3]. Due to the similarity of lesions, it is often difficult to distinguish melanoma and nevus in clinic. With the development of computer technology, more and more computer-aided melanoma detection techniques have been proposed and achieved good results.

Computer-aided melanoma detection is mainly categorised into two types, the first is based on machine learning. The main idea of this kind of method is to extract features and make classification using machine learning models. To be more specific, using some computer vision techniques to extract texture, colour, edge features from the image, representing the image information as a set of real numbers thus forming a feature vector. based on this feature vector, a suitable machine learning model is used to make the final classification. So the most important part of this method is feature extraction, because the performance of the classifier depends heavily on whether the process of feature extraction is successful or not.

The second method is based on deep learning, the idea of this method is to directly let a deep neural network receive the dermatological images, and perform pattern recognition, feature extraction itself, then finally output the classification results. This system is a kind of "black-box" model, because one does not really know what exactly is happening inside the network. The performance of this model mainly relies on whether the network structure is scientifically designed to fully extract the features in the image.

In this work, we propose a machine learning based approach that takes into account both the patient's personal information and the lesion location image information. Firstly, the lesion area is segmented using image segmentation techniques, and then texture and colour features are extracted from the original image and the image of the lesion area and represented as a feature vector. This vector is combined with the patient's personal information to form a super feature vector, and finally use XGBoost as the final classifier to classify the lesion type based on this information.

Our research work aims to achieve higher accuracy in identifying and classifying skin cancers by taking into account both the patient's personal information and the image information of the lesion location, contributing to the field by:

- Validate the rationale and necessity of considering patients' personal information to classify dermatological conditions.
- Identify key features for classifying melanoma and nevus.
- Find an appropriate combination of features to achieve the highest classification accuracy.

Our article will be organised in the following structure. Section 3 contains the specific background of melanoma and the current status of the development of computer-assisted melanoma methods and some of the landmark work. In section 4, the motivation of the idea of this paper and its rationale is presented .In section 5 the methodology proposed in this paper will be presented in detail. Section 6 will contain the EDA of the extracted features. Section 7 is a presentation of some machine learning methods and classification evaluation metrics. In section 8 and 9 are the experiment and evaluation of the classification methods proposed in this paper.

2 Motivation

The current diagnosis of melanoma is based on visual diagnosis and pathological examination by a doctor. Computer technology plays a very important role in visual diagnosis. The current computer-aided diagnosis of melanoma focuses on images of the lesion sites and analyses the images to make a diagnosis using computer vision technology. However, in clinical practice, doctors will also ask patients for personal information to help with the diagnosis.

Evidence suggests that Australia and New Zealand have the highest melanoma incidence rates in the world. Incidence rates are also high in Europe and North America, but lower in Asia, Africa and Latin America [4]. The incidence is also higher in men than in women[5]. Between the ages of 15 and 39, men are 55% more likely to die from melanoma than women in the same age group[6]. Demographic characteristics such as age and race are also very important in the diagnosis of this disease. In addition, the diagnosis of melanoma is also highly dependent on the location of the lesion, for example, melanoma is most commonly found on the legs in female patients and on the back in male patients[7]. In conclusion, automated diagnostic systems should also take into account the patient's personal information to make more accurate and credible predictions, rather than only focusing on analysing images of lesions.

The automatic diagnostic system proposed in this paper analyses the lesion location image while incorporating personal patient information such as gender, age, medical history and lesion location into the judgement in an attempt to achieve more accurate and robust prediction results.

3 Data Retrieving and Processing

This section discusses the datasets used in the study, the components of the datasets, and the exploratory data analyses of tabular data.

3.1 Dataset

Skin lesion images and patient data in Argentina The dataset used for the classification in this study consists of two components: (1) a collection of 1,616 images of skin lesions (1,270 contact-polarized dermoscopy images and 346 clinical images). (2) a tabular dataset containing 623 records of distinct patient information, including age, sex, personal history, family history, skin type, lesion's location.

Those image data and tabular data was obtained from a skin lesion dataset collected in Argentina[8]. This dataset is publicly available and was originally collected for research aims. In order to satisfy HIPAA[9] regulations, all personally identifiable information such as names and facial images, including tattoos or other potential identifiers, were removed or crop to protect individual privacy.

ISIC 2019 Challenge Datasets ISIC dataset is one of the most authoritative and famous skin lesion datasets in the world. It is organised and published by the International Skin Imaging Collaboration (ISIC) to promote research in the field of automatic diagnosis of skin diseases. This ISIC 2019 dataset is a dataset from a challenge published by the ISIC organisation in 2019. The challenge requires researchers to classify skin lesions with the highest possible accuracy. In this paper, this dataset was not employed for the purpose of classification. Instead, it was exclusively for transfer learning on the U-Net model for the segmentation task[10][11][12].

3.2 Data Cleaning

For the purposes of this study, which focuses on detecting melanoma and nevus skin lesions, we filtered the tabular dataset to include only relevant records. Initially, the dataset contained 10 different types of skin lesions, including Basal Cell Carcinoma, Melanoma, Squamous Cell Carcinoma, and others. After filtering, the dataset was reduced to 855 records, with 253 instances of melanoma and 602 instances of benign nevus lesions.

To ensure the highest possible image data quality for our task, another round of manual screening and selection was conducted. In this process, we carefully reviewed each image and removed images that were unsuitable for our analyses due to factors such as low resolution, obstructed views, or incomplete lesion areas. In the end, 152 melanoma images and 463 nevus images were selected, which were considered to be of sufficient quality for this study.

3.3 Missing Value Handling for Tabular Data

As illustrated in Figure 1. The 'age' variable has only 0.001% missing values, while 'family history' and 'personal history' have 34% and 25% missing values respectively. 'Skin type' has 0.04% missing values. The percentage of missing values ranged from a very low 0.001% to a high of 34%.

Considering the extremely low missing rates for age and skin type (0.001% and 0.04% respectively), these two variables are dropped from the analysis in order to avoid potential noise caused by such a small number of missing values. For the family history and personal history variables, which had substantial missing rates of 34% and 25%, a new category "Unreported" is created to represent the missing values. We hypothesise that patients may not have filled out their family history and personal history information because they were unaware or unwilling to disclose this information, resulting in higher missing rates for these two variables.

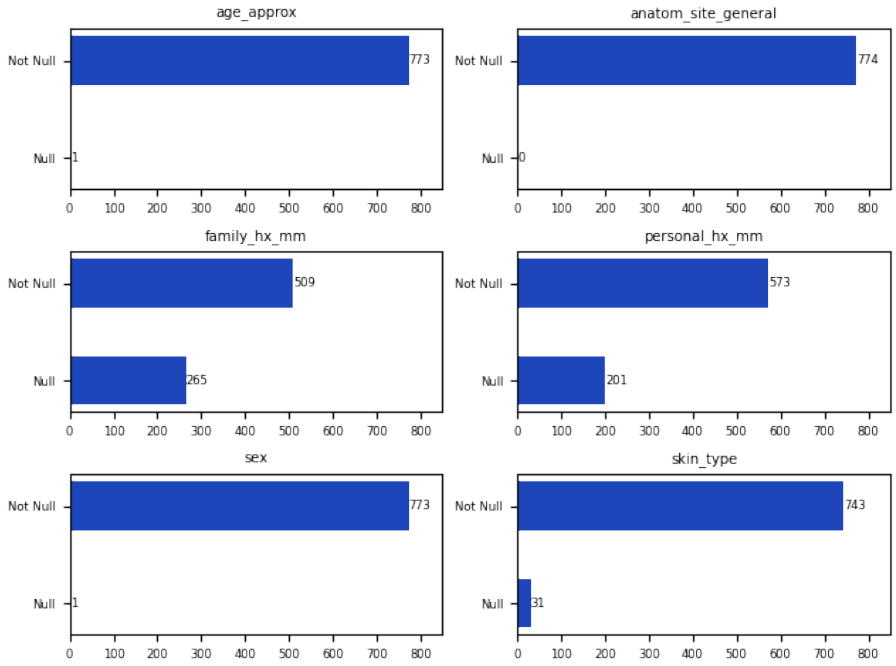

Figure 1: The distribution of missing tabular values across different variables

4 Background

4.1 Melanoma

Melanoma is a type of skin cancer develops from melanocytes, the pigment-producing cells in the skin. As is often the case, it can spread to other parts of the body, thus being considered the most dangerous form of skin cancer. It can also be life-threatening if not detected and treated early.

Worldwide, about 232100 (counts 1.7%) cases of all newly diagnosed primary malignant cancers (excluding non-melanoma skin cancer) are cases of cutaneous melanoma, and about 55500 cancer deaths (counts 0.7% of all cancer deaths) are due to cutaneous melanoma annually.[13]

Although melanoma can occur anywhere on the body, it most commonly develops on areas exposed to the sun, such as the face, neck, arms, and legs. Typically appears as an abnormal mole or spot on the skin, melanoma may exhibit changes in colour, size, shape, or texture over time. And this characteristic particularly adds to its complexity in medical detection.

4.2 Computer-aided melanoma detection

Deep learning has dramatically changed the field of dermatological image analysis, including the detection and diagnosis of melanoma, a deadly form of skin cancer. Recently, Convolution Neural Networks have achieved very good results in pattern recognition and segmentation[14].They excel at automatically extracting hierarchical features from images without the need for manual feature extraction. These models are trained on thousands of dermoscopic images to learn to distinguish between benign lesions and malignant melanomas, or to segment the skin lesion area.

Marriam et al[15]. proposed a skin lesion segmentation algorithm using a specially designed CNN called faster region-based convolution neural networks(RCNN) and fuzzy k-means clustering (FKM), achieving a competitive accuracy on several datasets. Pre-trained convolution neural network architectures[16] and hand-crafted methods[17], based on a transfer learning approach were used to classify skin lesions. Additionally, Khouloud et al.[18] developed a new approach based on pre-trained deep architectures.

There have also been some studies using machine learning techniques to classify skin lesions. Seeja et al.[19] developed a hybrid model based on the segmentation and classification of melanoma using dermoscopy images. Firstly, U-Net architecture was used in the segmentation process. Later, color, texture, and shape features based on hand-crafted methods were extracted using segmented images. Finally, combined features were fed into the machine learning classifiers (SVM, K-Nearest Neighbor, and Naïve Bayes) methods for melanoma classification. Khan etal.[20] used improved K-mean clustering to segment lesion area, followed by the calculation of texture and color features. Based on these features, SVM is utilized to make the final classification.

5 Methodology

In this section we discuss the components of our proposed method. The data obtained from the dataset is first pre-processed to enhance its quality. Texture and colour features are then extracted from the image data and combined with the original tabular data. Finally, all the information is used in a classifier to diagnose melanoma and nevus. Figure 1 presents the overview of the proposed methodology.

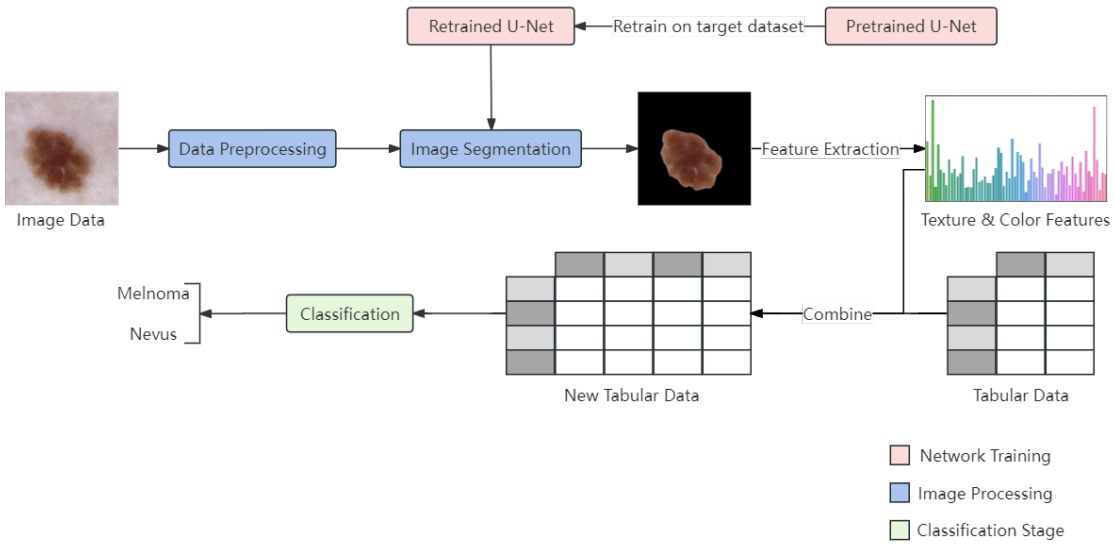

Figure 2: Overview of the proposed methodology

5.1 Pre-processing

Three pre-processing steps were conducted to prepare the images before inputting them into the U-net network and executing the feature extraction procedure.

5.1.1 Centre crop and resizing

In contrast to simply resizing the entire image, Samira (2022) presented a more effective approach, The process involved first cropping the central region of the image and resizing the cropped area to a lower resolution.

Following Samira's approach, all the images are automatically cropped around the centre and resized the resolution to 256x256, as shown in Figure 2. This method not only focuses in on the lesion region for future identification of feature, but crucially provides consistency across the dataset.

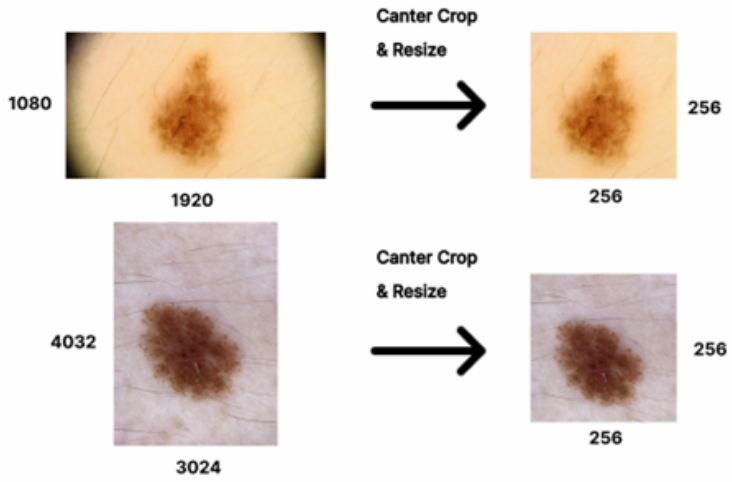

Figure 3: Crop centre and resize to a lower resolution

5.1.2 Hair removal

Artifacts such as hairs and bubbles in the image affect the accuracy of segmentation of skin lesion areas[21], thus many algorithms for removing hairs from the image have been proposed as pre-processing methods. Both morphology-based methods and Dullrazor are effective in removing hair artefacts and improving overall image quality[22]. Morphology-based methods use the Black Hat transform to enhance the hair region and then use threshold to segment the hair. DullRazor uses a combination of pixel comparison and adaptive filtering, which allows to deal better with images in the dataset that are of varying quality. In addition, DullRazor is less complex and faster to compute than morphological-based methods. Therefore, in this work, DullRazor is used for the hair removal process to extract and compute desired features from images more efficiently and accurately.

5.1.3 Data Augmentation

The field of bio-informatics is often constrained by the limited availability of large, high-quality datasets. This is mainly due to the private and sensitive content of medical imaging data. Medical data, especially visual information like skin images, often contain patient personal health information that is protected by strict privacy regulations. Therefore, researchers in the field of skin lesion analysis frequently face difficulties in obtaining large and diverse datasets to train and validate their machine learning model.

In this study, data augmentation techniques is employed to expand and increase the quality of the medical image dataset used for skin lesion analysis. The original skin lesion image transformed by using geometric transformations such as flipping and rotating them, as well as colour-based adjustments such as contrast enhancement.

Colour Augmentation For colour augmentation, adjustments have been made to the brightness and contrast of the image of the skin lesion. By amplifying the contrast helps to accentuate the distinction between the lesion area and the surrounding healthy skin areas. It also helps to make the borders and the texture of the lesion more visible.

Geometric Augmentation In addition to color augmentation, geometric augmentation techniques were also applied to further increase the training data samples. In this study, two geometric data augmentation techniques, reflection and rotation, were employed. Rotation is achieved by turning the image in a clockwise or counterclockwise direction over a specified range of 0 to 360 degrees, thereby increasing

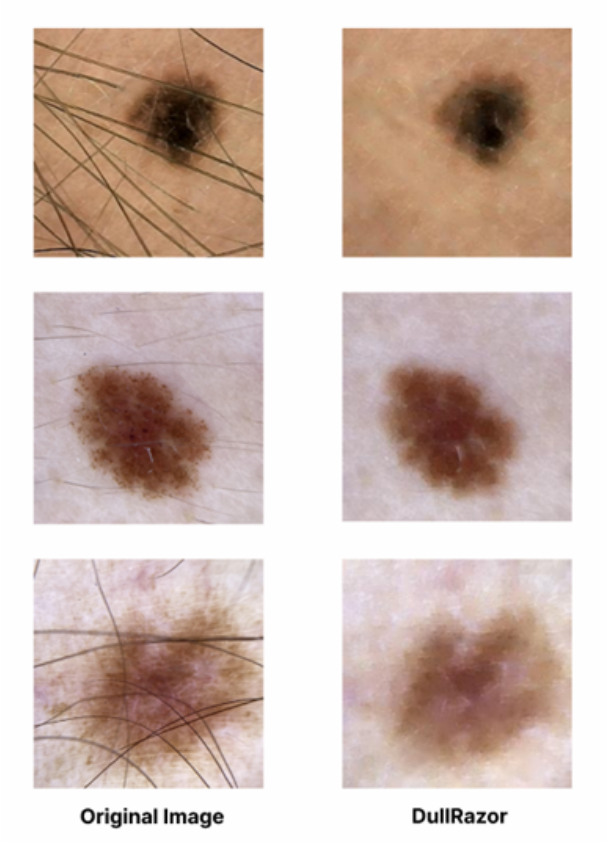

Figure 4: Images before and after applying Dullrazor

the size of the training data. For reflection, also known as flipping, three distinct types of flipping were applied: horizontal flipping, vertical flipping, and a combination of both horizontal and vertical flipping. Horizontal flipping generates a left-right mirror image of the original sample, while vertical flipping creates a top-bottom mirror image. Additionally, a combination of horizontal and vertical flipping were employed, resulting in a diagonal mirror image.

5.2 Image Segmentation

U-Net is a supervised deep learning model specialised in biomedical image segmentation. It is an improvement and development of the Fuzzy Cognitive Network(FCN) framework[7], which allows more accurate segmentation models to be trained with fewer images. In this work, U-Net is used to perform the image segmentation task.

Proposed method uses transfer learning to solve the problem of insufficient training data and domain shift.

The Argentina skin lesion dataset used for this work contains 764 images, after removing some of the blurred and too dark images, 615 images are left, this number is obviously not enough for deep learning based image segmentation techniques. Therefore, 4000 images were selected from the ISIC2019 dataset for initial training to obtain a pre-trained U-Net model. Then, 157 images were selected from the Argentina skin lesion dataset for repeated training. For these images, data enhancement operations such as adjusting brightness, increasing contrast, and image flipping are also performed to further increase the accuracy of segmentation. Finally, 527 images were obtained for repeated training.

Despite the fact that both the ISIC2019 and the Argentina skin lesion datasets are both skin lesion datasets, there is a significant domain shift between these two datasets. For example, the former does

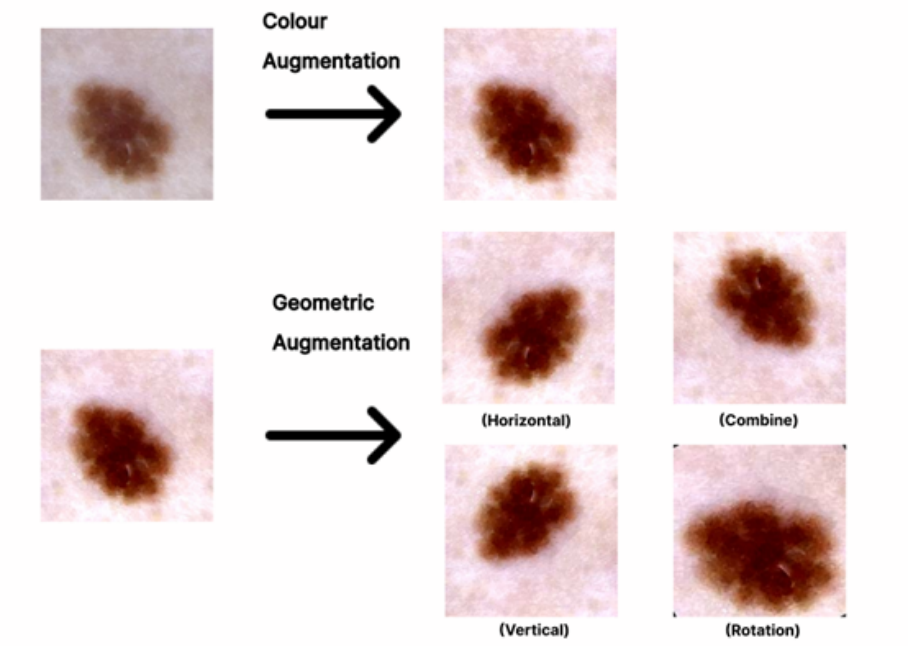

Figure 5: Data Augment by using Flipping and Adding Contrast

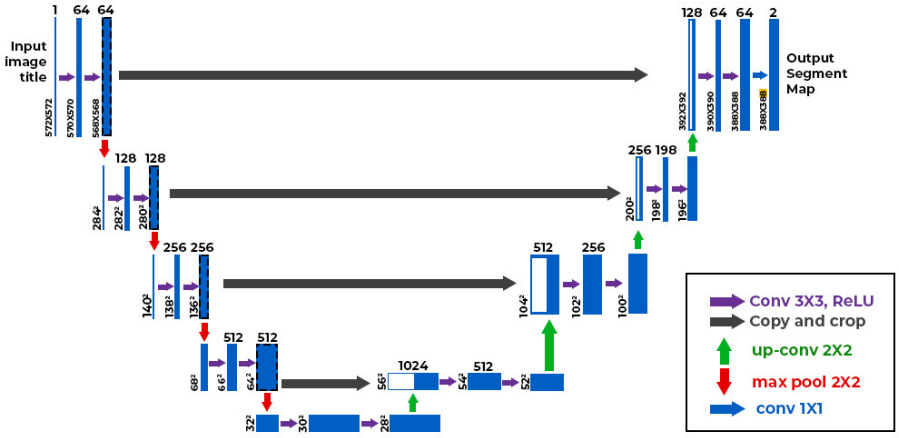

Figure 6: Network structure of U-Net

not indicate the ethnicity of the main source of the images, whereas in the latter all images were taken from Argentina. In the former, all the images were captured with a dermoscope, while in the latter, some images were captured with a dermoscope and the rest were captured with a smartphone. The experiments showed that the pre-trained U-Net model in ISIC2019 performed very poorly on the Argentina skin lesion datasets, so repeated training was performed to improve the performance of U-Net model on the target dataset.

5.3 Feature Extraction

Four feature sets were computed for the subsequent classification stage, including GLCM, LBP, C and Color, which is the Mean & Variance and Skewness of pixels in each channel of RGB. The texture and colour of skin lesions provides important clues about the nature of the lesion.

5.3.1 GLCM

Gray Level Co-occurrence Matrix (GLCM) analysis captures the statistical distribution of texture based on the relationships between pixel intensities, such as uniformity, contrast, and entropy. This method allows us to identify subtle texture differences within the lesion area, which may be crucial for distinguishing between benign and malignant lesions. To evaluate the textural characteristics of the image, the relation between both the reference pixel as well as the pixels around it is determined. For characterization of texture, it considers a set of characteristics extracted from uniform symmetrical directional GLCMs: energy, contrast, correlation, and homogeneity. Formulas for computing various GLCM functions are given in the following table.

Feature	Details	Formula
Contrast	Calculates the textural roughness among a pixel and its neighbors throughout the whole picture.	$\sum_{i=1}^{N}\sum_{j=1}^{N}(i-j)^2 \cdot P(i,j)$
Entropy	Measures randomness in the GLCM.	$-\sum_{i=1}^{N}\sum_{j=1}^{N}P(i,j)\log(P(i,j))$
Energy	Quantifies the level of disturbance or non-homogeneity.	$\sum_{i=1}^{N}\sum_{j=1}^{N}P(i,j)^2$
Correlation	Assesses the textural homogeneity in between 0 and 1.	$\frac{\sum_{i=1}^{N}\sum_{j=1}^{N}(i-\mu)(j-\mu)P(i,j)}{\sigma^2}$
Homogeneity	Relates to the impact of a pixel's surroundings on the entire picture.	$\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{P(i,j)}{1+(i-j)^2}$
Autocorrelation	Measures fineness and coarseness of texture.	$\sum_{i=1}^{N}\sum_{j=1}^{N}i \cdot j \cdot P(i,j)$

Figure 7: Calculation details of GLCM

In these formulas, N denotes the number of gray levels in the image, and P(i,j) is the normalized value of the GLCM at cell (i,j), representing the relative frequency.

5.3.2 LBP

Local Binary Pattern (LBP) is a computational tool utilized for characterizing the texture of a surface. The regularity of texture can be assessed through the distribution structure of the LBP histogram. LBP features encode texture-based data, which are integral in applications such as categorization and identification. Originally developed for analyzing 2D structures, the fundamental concept of this algorithm involves summarizing the local attributes of an image by comparing each pixel with its surrounding neighbors. Specifically, a central pixel is compared against its neighbors using a threshold value. If the central pixel's amplitude is greater than that of a neighbor, it is marked with a '1'; otherwise, it is marked with a '0'. This process generates an LBP operator that transforms the image into a vector of length N, enabling the analysis of the image's minute details.

5.3.3 C

'C feature' is derived from a subset of 'ABCD features'. Due to the inconsistent accuracy of 'ABD' features. 'C features' specifically refer to Color Variability, which is critical in identifying malignancies within lesions. Variations in color, such as multiple hues appearing within a single lesion, serve as crucial warning signs. To quantify this, the different colours present in the lesion (including tan, brown, black, red, white and blue) were identified to assess colour diversity. For each colour identified, one point was added to the overall colour score.Thus, a color scale ranging from 0 to 6 have been established, where each color observed adds one point to the score, thus allowing for a clear assessment of color variance in an image.

5.3.4 Color

Any colour distribution in an image can be expressed in terms of its moments, which is the mathematical basis of the method. And since the information about the colour distribution is mainly concentrated in the low-order moments, it is sufficient to express the colour distribution of an image using only the first-order (mean), second-order (variance) and third-order (skewness) moments of the colour. By incorporating these statistical moments into the feature computation, the representation of the colour distribution in an image can be improved, thereby improving the performance of subsequent machine learning algorithms.

In this work, the moments of the pixel values of the three RGB channels are computed separately to fine-tune the colour information.

$$average = (1/N)sum(P(i,j))$$
$$variance = sqrt(((1/N)sum(P(i,j)-u)^2),2)$$
$$skewness = sqrt(((1/N)sum(P(i,j)-u)^2),3)$$

where $P(i,j)$ denotes the pixel value of the correspond channel, u denotes the mean value of the pixel value of the correspond channel.

6 Exploratory Data Analysis

6.1 EDA for Tabular Data

To initially explore whether patients' personal information was associated with diagnostic outcomes, chi-square test of patient features versus diagnosis is implemented.

Figure 8 shows the results of the chi-square test. It can be seen that the p-values for all the features except 'sex' are significantly less than the significance level of 0.05. Such low p-values provides strong statistical evidence that patient personal information is relevant to differentiating between melanoma and nevus cases.

Variable Name	Type	Description	Chi-square p_value
anatom_site_ general	Categorical	Location of the lesion	3.9e-06
family_hx_mm	Categorical	Family history of corresponding lesion	5.2e-04
personal_hx_mm	Categorical	Personal history of corresponding lesion	5.3e-62
age_approx	Numerical	Approximate age of the lesion owner	3.0e-10
sex	Categorical	Patient gender	3.2e-02
skin_type	Categorical	Fitzpatrick scale skin type	2.9e-07

Figure 8: Chi-Square Test for Tabular Data

Figure 9 shows that of all the variables, age and personal medical history had the greatest difference in the distribution of melanoma and nevus cases.

Specifically, it is clear from the distribution of the age variable that most of the nevus patients were between 30-40 years of age, whereas the melanoma patients were between 75-85 years of age. Also, in terms of personal medical history, it was found that a significantly higher percentage of cases diagnosed with melanoma had a history of melanoma. This finding is particularly important because these two variables are likely to play a key role in the classification of melanoma and nevus cases.

Figure 9: Feature Distribution of Tabular Data

6.2 EDA for Features Extracted from Image Data

This section shows an exploratory analysis of the data extracted from the image, from which it is easy to initially identify some of the features will play an important role in the final classification stage.

GLCM It is clear from Figure 10 that for most of the features from the GLCM, both melanoma and nevus have very different distributions, especially for the ASM.The distribution of ASM values is more concentrated for melanoma and more dispersed for nevus.

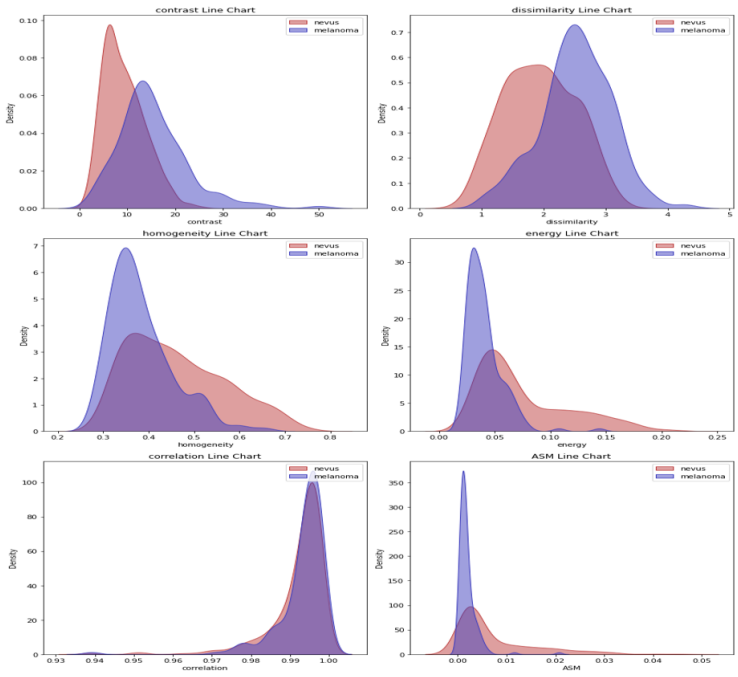

Figure 10: EDA for GLCM

LBP LBP is another set of features to extract the texture information of the image, and it can be seen from Figure 11 that the three LBP features computed in this work all have very different distributions.

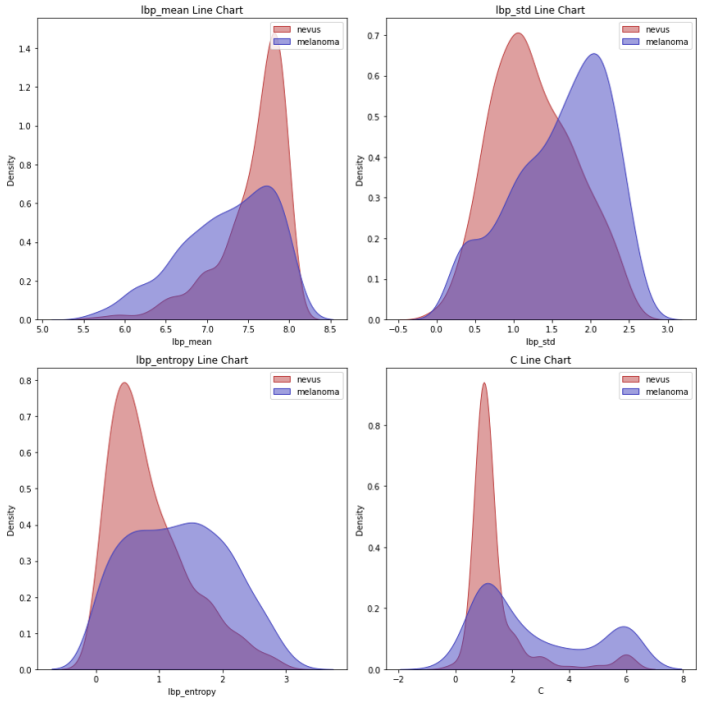

Figure 11: EDA for LBP and C

C and Color Both C and Color are features for extracting colour information.C can represent the colour inhomogeneity within the lesion area and Color can measure the colour distribution in the lesion area at the pixel level. It can be seen from Figure 12 that melanoma tends to have a higher value of C, indicating that it has heterogeneous colours within it. In addition, the pixel value statistics for the R and B channels are very different because melanoma area may have unique colours such as red and blue.

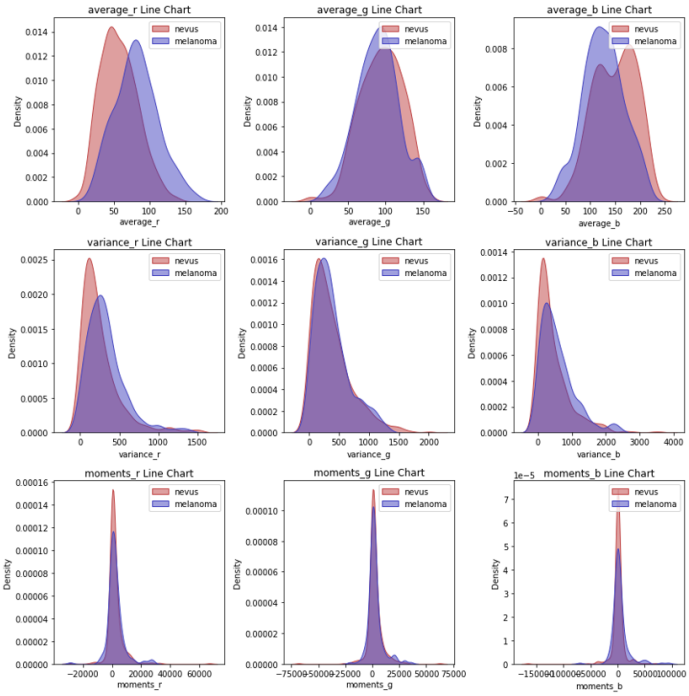

Figure 12: EDA for Color

7 Result and Discussion

7.1 Comparative feature combination analysis

Multiple feature combinations were used in the experiments. A comparative analysis has been made in Figure 13 and Figure 14. It can be seen that the classifier achieved an accuracy of 86% when using only tabular data. There is almost no difference in classification accuracy when using only GLCM and when using GLCM+LBP, with a small increase in f1-score. This may be because both GLCM and LBP represent texture information of the lesion area, so the combination of these two features does not bring more valuable information to the model. C and Color are colour features that represent the lesion area, and due to the nature of the features, Color contains more information than C. Thus, a more significant improvement in accuracy and f1-score is obtained when both C and Colour features are introduced into the model. When all the features extracted from the images were combined with the tabular data, the model obtained the highest accuracy and the highest f1 score. Thus, both the images and the tabular data contain meaningful information for the classification of skin lesions. Notably, the classifier achieved the lowest accuracy but the highest f1-score when using only the lesion images classified directly by the convolution neural network VGG16.

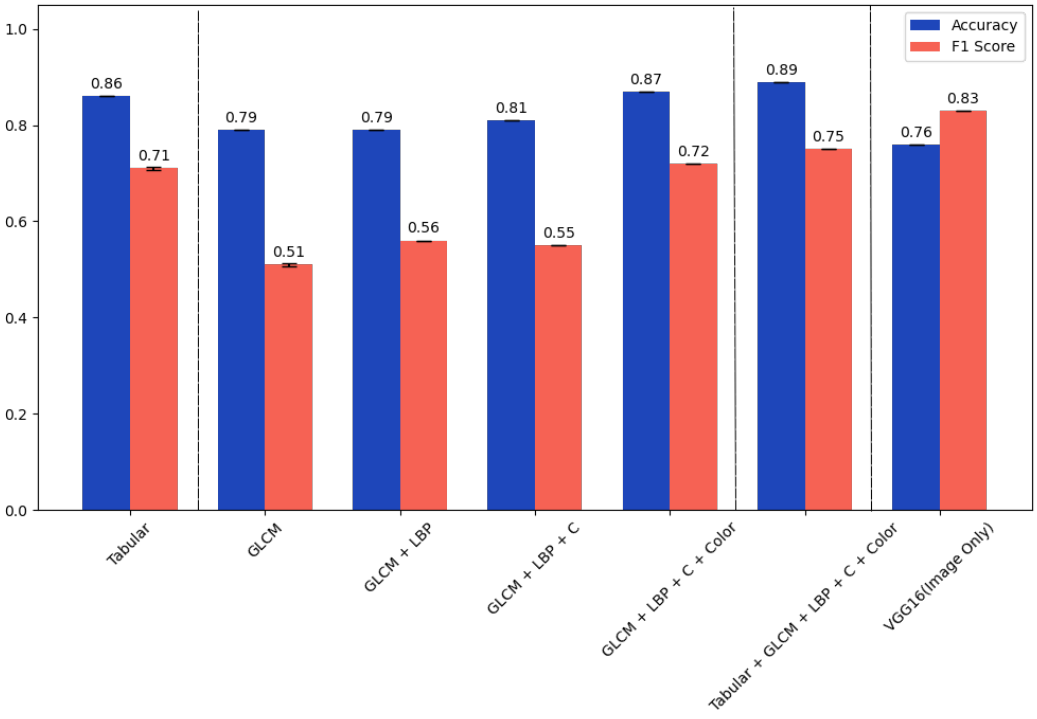

Figure 13: Comparison of Accuracy and F1 Score with Variance

In terms of AUC, if only one set of features is used, GLCM can give the best result and the AUC can reach 0.8, which is not very high. There is a small increase in AUC when the two features representing texture information, GLCM and LBP, are used in combination, but this increase is not significant. If the features C and Color which represent colour information are further introduced into the model, the AUC increases to 0.9. On the other hand, if only the tabular data is used for classification, the AUC also reaches 0.92, indicating that a good classification can be achieved by focusing only on the patient's personal information. When combining all available information, the AUC reached a maximum of 0.95, indicating that information from different sources is valuable for the skin lesion classification task. It is also necessary to focus on other sources of information, such as the patient's personal information, while trying to extract sufficiently complete information from the image data to achieve higher classification accuracy.

7.2 Feature importance analysis

In order to explore which features are of relatively high importance for melanoma identification and to provide more interpretative results for automatic skin lesion classification, two methods are used in the process of feature importance calculation, which are XGBoost feature importance and permutation importance.

The results are similar for both calculations. In terms of patients' personal information, the four most important characteristics are all patients' personal medical history, family medical history and age. This matches well with the patient's situation that doctors focus on when making a diagnosis[23], suggesting that the proposed method can mimic clinical dermatologists in making a diagnosis to some extent. The features extracted from the images also play an important role. Among the feature importance calculated by both methods, the two most important image features are ASM and average_r. ASM is a texture feature from GLCM. An important basis for the diagnosis of melanoma is the inhomogeneity of the internal texture and the uneven distribution of colours[24], which means the texture in the skin lesion is

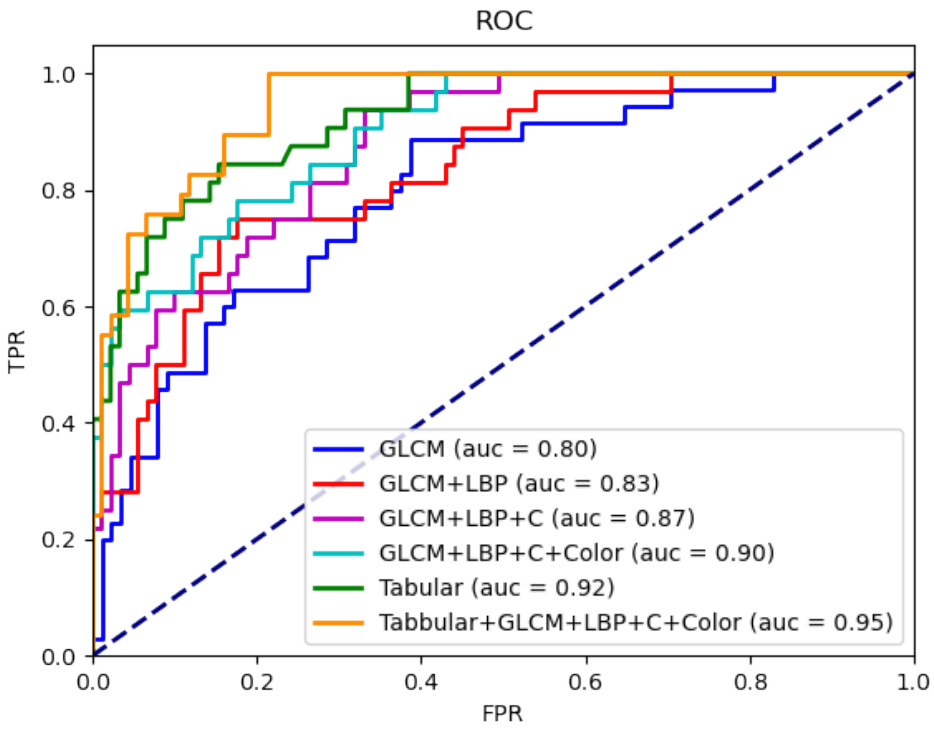

Figure 14: ROC curve of Binary Classification for Different Feature Combination

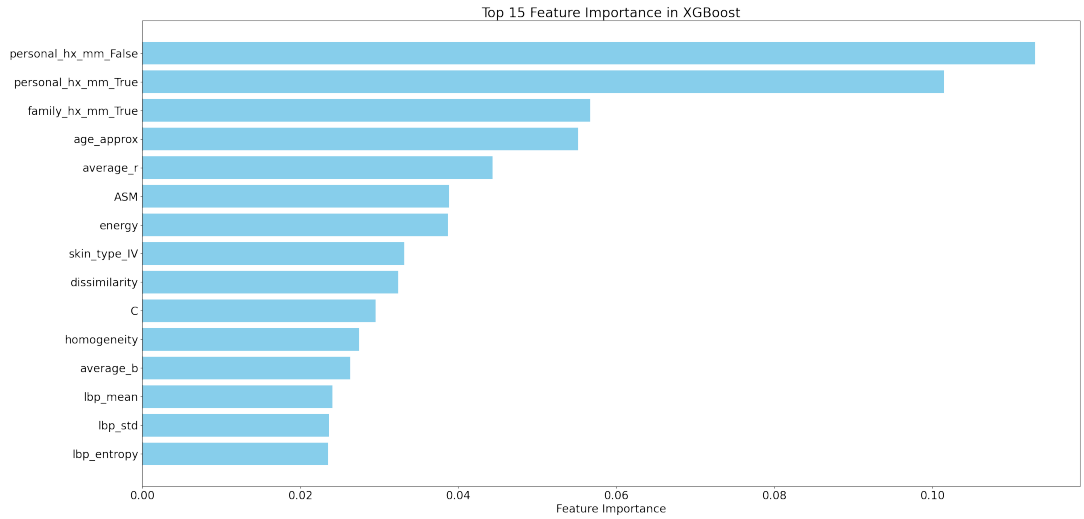

Figure 15: Top 15 Important Features for XGBoost

uneven. so ASM is an important feature for classifying the difference between melanoma and nevus. In addition, due to the malignant transformation of melanocytes, red, black and blue colour changes occur inside the melanoma area[24], so for melanoma, the average value of the red channel pixel value will be larger than that of the nevus, which means that the value of average_r will be larger.

Weight	Feature
0.0374 ± 0.0166	age_approx
0.0276 ± 0.0166	personal_hx_mm_True
0.0276 ± 0.0166	personal_hx_mm_False
0.0260 ± 0.0315	average_r
0.0211 ± 0.0166	ASM
0.0179 ± 0.0159	family_hx_mm_Unreported
0.0146 ± 0.0122	variance_g
0.0114 ± 0.0080	homogeneity
0.0114 ± 0.0166	correlation
0.0098 ± 0.0159	anatom_site_general_posterior torso
0.0098 ± 0.0280	moments_b
0.0081 ± 0.0000	anatom_site_general_head/neck
0.0081 ± 0.0103	anatom_site_general_lower extremity
0.0081 ± 0.0103	contrast
0.0049 ± 0.0080	C
0.0033 ± 0.0264	variance_r
0.0033 ± 0.0166	lbp_mean
0.0033 ± 0.0130	lbp_std
0.0033 ± 0.0130	energy
0.0033 ± 0.0166	average_g

Figure 16: Permutation Feature Importance

8 Conclusion

In this research, we presented an automatic diagnosis system for melanoma and nevus.One of the most exciting aspects of this system is that it not only focuses on the image data, but also takes into account the patient's personal information to perform the classification. Experimental results showed that on a dermatological dataset collected in Argentina, this operation was effective in improving the classification accuracy to 89%, increasing the AUC value by 0.95.

It was found that there is a large domain transformation between different dermatological datasets. The U-net trained on ISIC2019 could not achieve good segmentation results on the target dataset, so the transfer learning strategy had to be used. In addition, the system identified features such as age, patient history, ASM, avarage_r, etc., which are very crucial for classification, providing suggestions and guidelines for future research.

9 Group Work

References

[1] C. D. S. A. F. Jerant, J. T. Johnson and T. J. Caffrey, "Early detection and treatment of skin cancer," *American Family Physician*, vol. 32, no. 2, pp. 357–368, 2000.

[2] R. Kasmi and K. Mokrani, "Classification of malignant melanoma and benign skin lesions: Implementation of automatic abcd rule," *IET Image Process*, vol. 10, no. 6, pp. 448–455, 2016.

[3] D. S. R. R. J. Friedman and A. W. Kopf, "Early detection of malignant melanoma: The role of physician examination and self-examination of the skin," *CA, Cancer J. Clinician*, vol. 35, no. 3, pp. 130–151, 1985.

[4] C. P. Wild, B. W. Stewart, and C. Wild, *World cancer report 2014*. World Health Organization Geneva, Switzerland, 2014.

[5] D. E. Fisher and A. C. Geller, "Disproportionate burden of melanoma mortality in young us men: the possible role of biology and behavior," *JAMA dermatology*, vol. 149, no. 8, pp. 903–904, 2013.

[6] K. G. Paulson, D. Gupta, T. S. Kim, J. R. Veatch, D. R. Byrd, S. Bhatia, K. Wojcik, A. G. Chapuis, J. A. Thompson, M. M. Madeleine *et al.*, "Age-specific incidence of melanoma in the united states," *JAMA dermatology*, vol. 156, no. 1, pp. 57–64, 2020.

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer, 2015, pp. 234–241.

[8] M. A. Ricci Lara, M. V. Rodríguez Kowalczuk, M. Lisa Eliceche, M. G. Ferraresso, D. R. Luna, S. E. Benitez, and L. D. Mazzuoccolo, "A dataset of skin lesion images collected in argentina for the evaluation of ai tools in this population," *Scientific Data*, vol. 10, no. 1, p. 712, 2023.

[9] "Hipaa. health information privacy," https://www.hhs.gov/hipaa/index.html.

[10] N. C. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler *et al.*, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic)," in *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*. IEEE, 2018, pp. 168–172.

[11] M. Combalia, N. C. Codella, V. Rotemberg, B. Helba, V. Vilaplana, O. Reiter, C. Carrera, A. Barreiro, A. C. Halpern, S. Puig *et al.*, "Bcn20000: Dermoscopic lesions in the wild," *arXiv preprint arXiv:1908.02288*, 2019.

[12] P. Tschandl, C. Rosendahl, and H. Kittler, "The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific data*, vol. 5, no. 1, pp. 1–9, 2018.

[13] D. Schadendorf, A. C. Van Akkooi, C. Berking, K. G. Griewank, R. Gutzmer, A. Hauschild, A. Stang, A. Roesch, and S. Ugurel, "Melanoma," *The Lancet*, vol. 392, no. 10151, pp. 971–984, 2018.

[14] F. Alenezi, A. Armghan, and K. Polat, "A multi-stage melanoma recognition framework with deep residual neural network and hyperparameter optimization-based decision support in dermoscopy images," *Expert Systems with Applications*, vol. 215, p. 119352, 2023.

[15] M. Nawaz, Z. Mehmood, T. Nazir, R. A. Naqvi, A. Rehman, M. Iqbal, and T. Saba, "Skin cancer detection from dermoscopic images using deep learning and fuzzy k-means clustering," *Microscopy research and technique*, vol. 85, no. 1, pp. 339–351, 2022.

[16] M. A. R. Ratul, M. H. Mozaffari, W.-S. Lee, and E. Parimbelli, "Skin lesions classification using deep learning based on dilated convolution," *BioRxiv*, p. 860700, 2019.

[17] J. Jaworek-Korjakowska, A. Brodzicki, B. Cassidy, C. Kendrick, and M. H. Yap, "Interpretability of a deep learning based approach for the classification of skin lesions into main anatomic body sites," *Cancers*, vol. 13, no. 23, p. 6048, 2021.

[18] S. Khouloud, M. Ahlem, T. Fadel, and S. Amel, "W-net and inception residual network for skin lesion segmentation and classification," *Applied Intelligence*, vol. 52, no. 4, pp. 3976–3994, 2022.

[19] R. Seeja and A. Suresh, "Deep learning based skin lesion segmentation and classification of melanoma using support vector machine (svm)," *Asian Pacific journal of cancer prevention: APJCP*, vol. 20, no. 5, p. 1555, 2019.

[20] M. Q. Khan, A. Hussain, S. U. Rehman, U. Khan, M. Maqsood, K. Mehmood, and M. A. Khan, "Classification of melanoma and nevus in digital images for diagnosis of skin cancer," *IEEE Access*, vol. 7, pp. 90 132–90 144, 2019.

[21] R. B. Oliveira, M. E. Filho, Z. Ma, J. P. Papa, A. S. Pereira, and J. M. R. Tavares, "Computational methods for the image segmentation of pigmented skin lesions: A review," *Computer Methods and Programs in Biomedicine*, vol. 131, pp. 127–141, 2016.

[22] S. Nazari and R. Garcia, "Automatic skin cancer detection using clinical images: A comprehensive review," *Life*, vol. 13, no. 11, p. 2123, 2023.

[23] E. R. Farmer, R. Gonin, and M. P. Hanna, "Discordance in the histopathologic diagnosis of melanoma and melanocytic nevi between expert pathologists," *Human pathology*, vol. 27, no. 6, pp. 528–531, 1996.

[24] V. Zeljkovic, C. Druzgalski, S. Bojic-Minic, C. Tameze, and P. Mayorga, "Supplemental melanoma diagnosis for darker skin complexion gradients," in *2015 Pan American Health Care Exchanges (PAHCE)*. IEEE, 2015, pp. 1–8.