Tech Science Press

Statistical Histogram Decision Based Contrast Categorization of Skin Lesion Datasets Dermoscopic Images

Rabia Javed[1,2], Mohd Shafry Mohd Rahim[1], Tanzila Saba[3], Suliman Mohamed Fati[3], Amjad Rehman[3,*] and Usman Tariq[4]

[1]School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia Johor Bahru, 81310, Malaysia
[2]Department of Computer Science, Lahore College for Women University, Lahore, 54000, Pakistan
[3]College of Computer and Information Sciences, Prince Sultan University, Riyadh, 11589, Saudi Arabia
[4]College of Computer Engineering and Science, Prince Sattam bin Abdulaziz University, Alkharj, Saudi Arabia
[*]Corresponding Author: Amjad Rehman. Email: rkamjad@gmail.com
Received: 08 October 2020; Accepted: 22 December 2020

Abstract: Most of the melanoma cases of skin cancer are the life-threatening form of cancer. It is prevalent among the Caucasian group of people due to their light skin tone. Melanoma is the second most common cancer that hits the age group of 15–29 years. The high number of cases has increased the importance of automated systems for diagnosing. The diagnosis should be fast and accurate for the early treatment of melanoma. It should remove the need for biopsies and provide stable diagnostic results. Automation requires large quantities of images. Skin lesion datasets contain various kinds of dermoscopic images for the detection of melanoma. Three publicly available benchmark skin lesion datasets, ISIC 2017, ISBI 2016, and PH2, are used for the experiments. Currently, the ISIC archive and PH2 are the most challenging and demanding dermoscopic datasets. These datasets' pre-analysis is necessary to overcome contrast variations, under or over segmented images boundary extraction, and accurate skin lesion classification. In this paper, we proposed the statistical histogram-based method for the pre-categorization of skin lesion datasets. The image histogram properties are utilized to check the image contrast variations and categorized these images into high and low contrast images. The two performance measures, processing time and efficiency, are computed for evaluation of the proposed method. Our results showed that the proposed methodology improves the pre-processing efficiency of 77% of ISIC 2017, 67% of ISBI 2016, and 92.5% of PH2 datasets.

Keywords: Cancer; healthcare; contrast enhancement; dermoscopic images; skin lesion; low contrast images; WHO

1 Introduction

Skin cancer is the most-known type of cancer. It is usually categorized into two classes: melanoma and non-melanoma. However, melanoma is one of the dangerous forms of skin cancer that can cause patient death. Melanoma is considered a lethal disease among the other common

skin cancer diseases like basal cell carcinoma and squamous cell carcinoma. This skin cancer is generally associated with overexposure to ultraviolet light. According to the statistics of the year 2020 [1], the estimated new cases of invasive melanoma are around 100,350 (about 60,190 in males and 40,160 in females) in the United States. The report also estimated that 6,850 people would die due to melanoma (nearly 4,610 in males and 2,240 in females). Early detection thus gives more significant chances of treatment. Hence the importance of melanoma detection is analyzed. The dermatologist has utilized several noninvasive imaging techniques like dermoscopy, confocal microscopy, and optical coherence tomography for skin lesion detection and classification [2,3].

Dermoscopy is one of the standards and effective imaging techniques used in the diagnosis of skin cancer. The dermoscopy captures the $20\times$ magnified lesion image, but sometimes due to the lighting effects, the image contrast is varying. These variations in image contrast create difficulty for skin lesion classification methods and reduce these method's accuracy. Computer-aided diagnosis (CAD) systems for skin lesion classification mostly have these fundamental steps, including image pre-processing, lesion segmentation, feature extraction, and lesion classification. Many researchers used the contrast enhancement technique to solve the contrast variation issue in the image pre-processing step before moving on to the segmentation and classification steps [4]. However, due to pre-processing on large un-categorized datasets [5], the computational time is increased, and also it reduces the efficiency of the method.

Moreover, some of the high contrast images already have high contrast pixel values. After applying the contrast enhancement techniques on these high contrast images, the segmentation method's accuracy will also be affected. Consequently, it is not necessary to use the image contrast enhancement on high contrast images. In this article, we proposed a novel statistical histogram-based decision method to address this problem. Our contributions are two-fold. Firstly, a statistical histogram decision method is implemented for the automated categorization of skin lesion datasets. The histogram upper bin and lower bin properties are used to separate the high and low contrast images. Secondly, the pre-processing histogram-based contrast stretching technique is only applied to the low contrast images. The significant contributions of this article are as follows:

- A novel approach proposed using image histogram properties to accomplish skin lesion datasets analysis before the pre-processing step.
- Formulate the statistical equations for separating the dermoscopic images into high/normal and low contrast images.
- The pre-processing step's efficiency is improved by gathering the histogram of images' metadata and enhancing only the contrast of low contrast images.

The rest of the paper is planned as follows. Section 2 contains some of the earlier work focused on dataset analysis and improvements in the pre-processing step. The details of the proposed method for skin lesion datasets categorization using the histogram-based formulation are explained in Section 3. The experimental results are discussed in Section 4. Finally, the conclusion with future work in Section 5.

2 Related Work

Every dataset plays a vital part in developing and validating every CAD system [6,7]. Understanding raw datasets are necessary for further processing like image segmentation, feature extraction, and classification. The skin lesion images comprise labeled images with metadata information, e.g., lesion type, personage, gender, etc. Many researchers used different augmentation

techniques like geometric transformation, color space transformation [8], cropping, rotation, translation, flipping, zooming, mixing image, blurring, and kernel filter [9] for dataset analysis. The outcome of the research performed in existing literature states that it is highly dependent on the number of images and their quality [10,11]. There is no established standard for capturing the skin lesion images. That is the reason the results of automated skin lesion diagnosis are observed to affect accuracy severely.

In the pre-processing step, some researchers solve the issue of artifacts and hair removal. Furthermore, some of the authors focus on increasing the contrast of images through contrast enhancement techniques [12,13]. Also, the reflection issue is solved by illumination correction and color space transformation [14]. It was discussed by Rahman et al. [15] that when images are acquired from different datasets using the different image capturing devices under changing environments such as lighting, lens, textures, backgrounds, and others made the identification and the classification tasks of lesions very difficult. The image histogram is used for many different purposes like pre-processing, image segmentation, feature extraction [16]. Different devices appear in the image due to the various devices, which makes the identification process more challenging [17]. The difficulty of isolating the lesion segment in the images increases if captured images contrast is low [18]. Abbas et al. [19] proposed enhancing lesion image contrast by adjusting and mapping the lesion pixels' intensity values in the specified range in CIE $L * a * b$ color space. The color histogram properties are utilized for the discrimination between benign and melanoma lesions [20]. The image histogram is used in the region growing method to segmentation pigmented skin lesions [21]. The authors [22] used the histogram for thresholding value for clustering segmentation of low contrast or different tone images. The histogram's luminance level is calculated to create a binary image that feeds to the segmentation stage.

The histogram is also utilized to extract global and local features [23,24]. The color histograms have been applied to classify the lesion into melanoma and benign [25]. The color histogram analysis is performed to extract the color characteristics of melanoma, and these acquired characteristics are used to differentiate between melanoma and benign lesions [26]. Moreover, the histogram techniques like histogram equalization, histogram stretching and histogram matching, etc., are also used for image normalization. These approaches' significant side effect is the over-exaggeration of the noise in the area, which has a comparatively low-intensity range. Contrast Limited Adaptive Histogram Equalization (CLAHE) is one proposed technique that might be applied to mitigate noise's above-mentioned side effects [27,28]. Many pre-processing procedures, like image augmentation and normalization, are used to balance and filter skin lesion dataset's images. The different types of contrast enhancement techniques are used for improving the quality of dermoscopic images [29,30]. Contrast stretching is one of the methods for image enhancement by stretching intensity values. Different contrast stretching types exist, such as local contrast, global contrast, partial contrast, bright contrast, and dark contrast stretching techniques. The local contrast stretching technique is specially used for low contrast. The global stretching is also used to enhance the image quality by taking global contrast of the image.

The most crucial problem in lesion detection and classification is the low contrast dermoscopic images. The low contrast images affect the segmentation performance due to the similarity between healthy skin and infected lesion areas. The second major issue is lesion boundary extraction due to lesions irregular and different shapes. The third issue is the accurate classification of the lesion into benign and melanoma because a wide range of color similarity exists among lesion types. Currently, in existing work, all of these issues are present because of un-categorized skin

lesion datasets. Before using the skin lesion datasets for lesion detection and classification of dermoscopic images, there should be methods to analyze these datasets.

3 Proposed Method

The proposed methodology is divided into two phases. In phase 1, the analysis of skin lesion datasets is performed, then an automated intelligent histogram-based method is implemented that categorized the dermoscopic images into high and low contrast images. In phase 2, the pre-processing is performed only on identified low contrast images. The proposed method is graphically represented in Fig. 1. The detail of each phase is explained below:
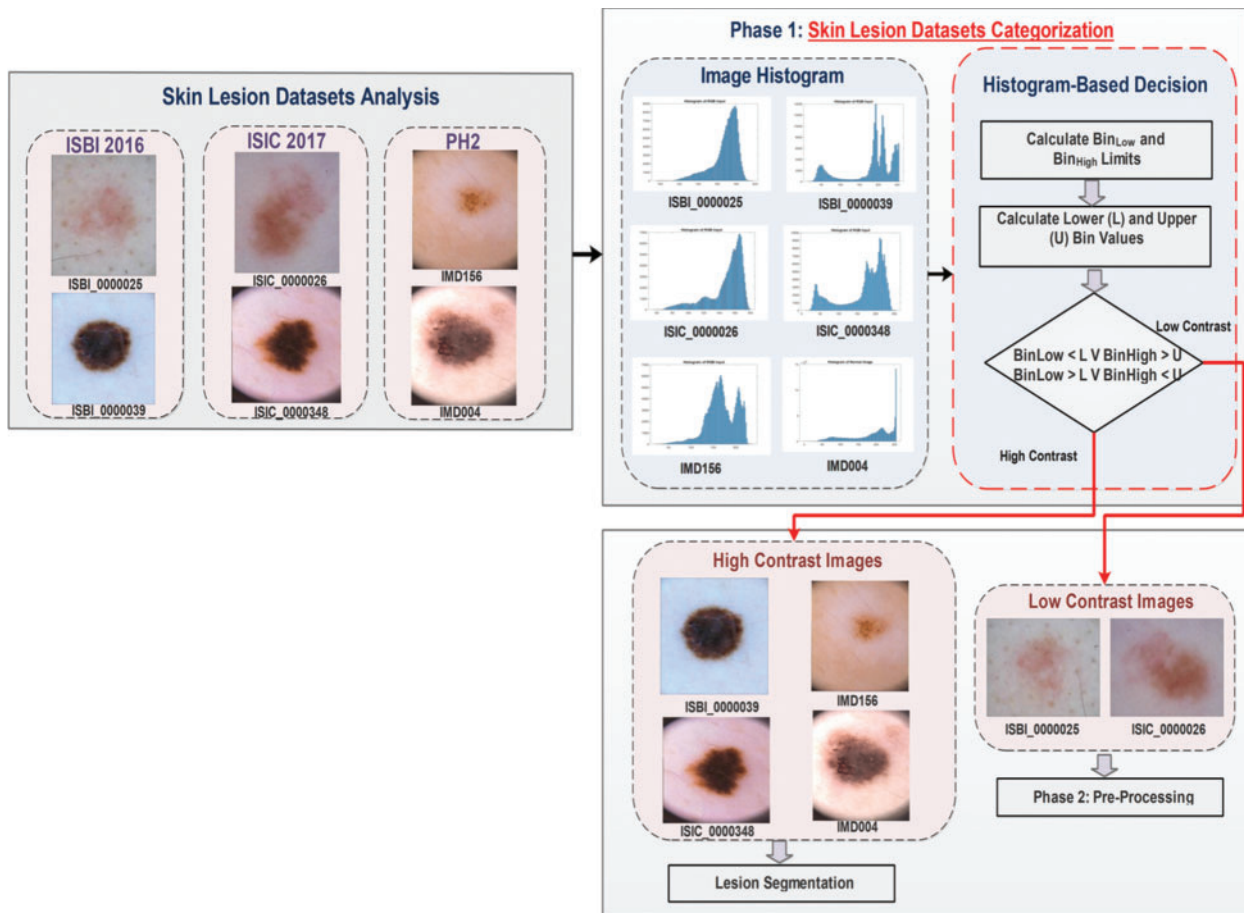

Figure 1: A proposed method for categorization into high and low contrast images

3.1 Skin Lesion Dataset Analysis

The skin lesion datasets ISBI (International Symposium on Biomedical Imaging) 2016, ISCI (International Skin Imaging Collaboration) 2017, and PH2 contain various dermoscopic images. Huge variations and challenges exist in these datasets. For this research, the skin lesion datasets images are categorized into three classes: (1) artifacts images, (2) variations in image contrast, and (3) variations in lesion properties. The first class (artifacts images) contains different types of

artifacts in images like (a) dark corner, (b) maker ink, (c) gel bubbles, (d) circle chart, (e) ruler marks, and (f) skin hairs. An artifact is anything rather than the lesion that is detected while capturing the lesion from dermoscopy. These artifacts are naturally not present in the skin. Some of them create a problem during the digital image processing steps that will be performed for lesion classification. These artifacts reduced the performance of the pre-processing step and also created problems in lesion boundary extraction. The second class (variations in image contrast) has two types of images: (a) high contrast images and (b) low contrast images.

The low contrast or quality of images shows that human eyes cannot easily differentiate between healthy/normal skin and cancerous/lesion skin. The third class (variations in lesion properties) is divided into (a) under segmented images and (b) over segmented images. The under segmented images mean that the lesion size is too small and difficult to analyze. The over segmented images are like pigmented skin where a lesion is spread all over the skin area. The contrast is defined as uniquely different in terms of shade or glows representing an object (in real or captured digitally). The same view's objects should have distinguishable features such as color or brightness labeled as high contrast images or objects. Some contrast enhancing techniques such as contrast stretching or normalization work by expanding or stretching the intensity values to the desired intensity values. The contrast of the image can be high or low. It depends on the tone range of the image. In a high contrast image, the full range of tone appears from bright to dark. The high contrast images contain normal images that are high in contrast. A high or well-contrasted image has gray levels (histogram) spread out over much of the range. A low contrasted image has gray levels (histogram) cluttered in the center.

3.2 Skin Lesion Datasets Categorization

The first phase of the proposed method is the categorization of images into high and low contrast. The purpose of categorizing images is to minimize the time consumed during the pre-processing step performed on the large datasets. The researched method is achieved in five steps, as illustrated in Fig. 2. Step 1 is to create the histogram for all images, followed by step 2 that calculates the histogram properties, while step 3 is to select the two histogram properties BinLow and BinHigh. The lower (L) and upper (U) limits of an image are calculated in step 4, followed by step 5 that decides whether an image is a high or low contrast based upon the formula.

3.2.1 Create a Histogram for All Images

A histogram is a visual representation of an image that presents the intensity values of pixels. These intensity values are plotted on the x-axis (vertical) by different intervals, also known as bars or bins. The histogram gives a detailed view of all pixel values of an image that helps inspect an image.

Also, the brightness and contrast of an image are easily identified by analyzing the distributed intensity values. The histogram decision creates the histogram for all the dermoscopic images presents in benchmark datasets ISIC 2017, ISBI 2016, and PH2. The histogram of the skin lesion image is generated and each bar defines a range of values between the minimum and maximum values.

3.2.2 Calculate Histogram Properties

Histogram properties, as define in MathWorks, control the appearance and behavior of the histogram. Some of the properties of the histogram are: (i) the number of bins (BinCounts), (ii) width of bins (BinWidth), (iii) edges of bins (BinEdges), and (iv) bin limits (BinLimits). The bin limits are further comprised of two values, which are bin low and bin high limit. For the

analysis of skin lesion images, the histogram properties/features such as bin count, bin edges, bin low value, bin high value, and bin width are calculated. Here, the first ten images detail results are shown in Tab. 1.
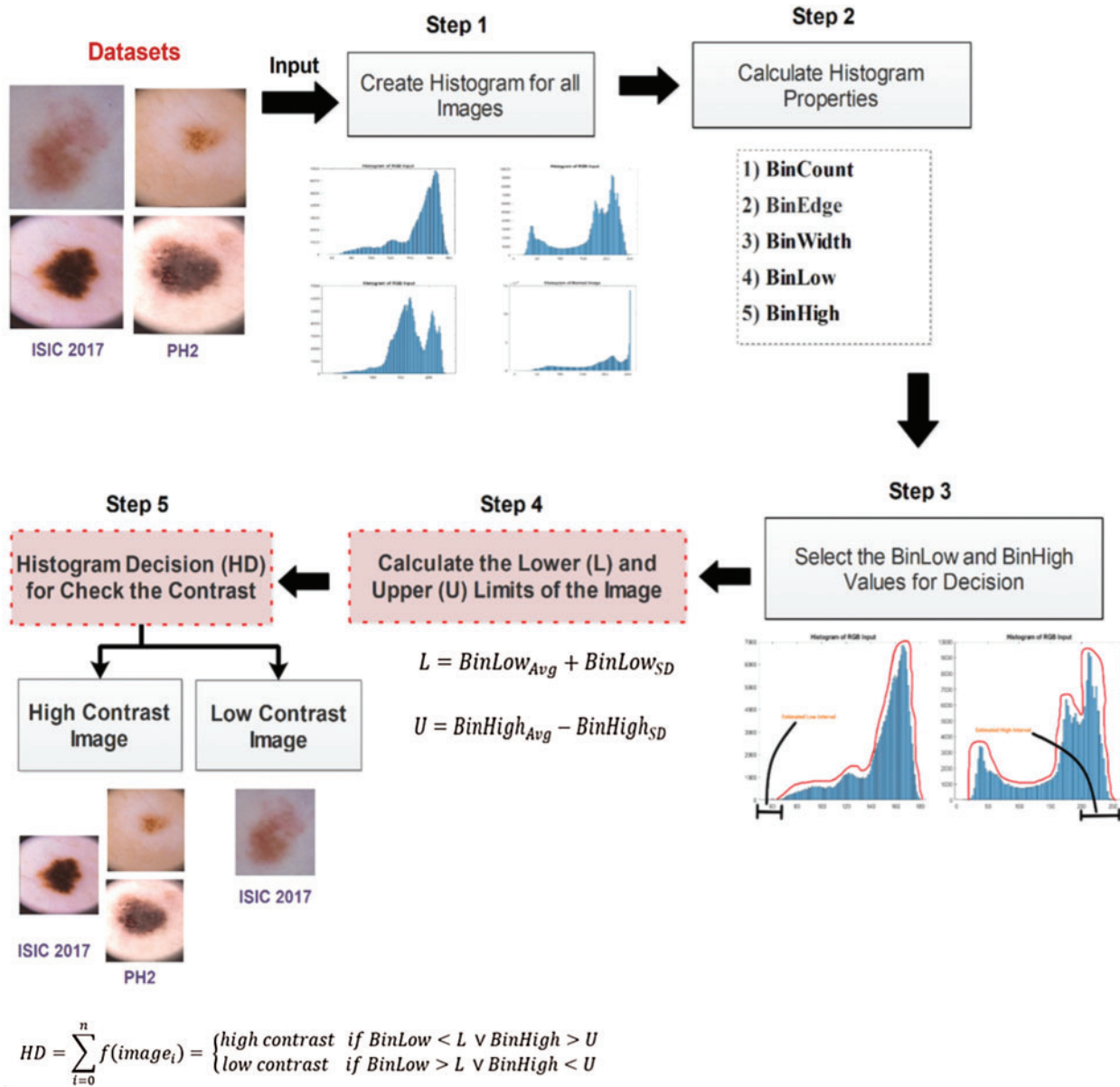

Figure 2: Process of histogram-based decision method

The same procedure is followed for the datasets ISBI 2016 and PH2. Firstly, the histogram is created for all the images; then, the histogram properties are calculated. The detailed results of the first ten images are shown in Tab. 2.

The complete histogram properties are also calculated for the PH2 dataset, as shown in Tab. 3.

Table 1: Histogram properties for ISIC 2017 dataset images

Image name	Histogram properties				
	BinCounts	BinEdges	BinLow	BinHigh	BinWidth
ISIC_0000000	47	48	20	255	5
ISIC_0000001	109	110	10	228	2
ISIC_0000002	85	86	0	255	3
ISIC_0000004	51	52	0	255	5
ISIC_0000006	118	119	0	236	2
ISIC_0000007	83	84	0	249	3
ISIC_0000008	79	80	12	249	3
ISIC_0000009	112	113	0	224	2
ISIC_0000010	203	204	6	209	1
ISIC_0000011	74	75	0	222	3

Table 2: Histogram properties for ISBI 2016 dataset images
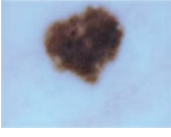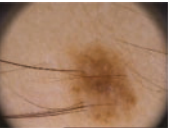
Image name	Histogram properties				
	BinCounts	BinEdges	BinLow	BinHigh	BinWidth
1002	160	161	55	215	215
1003	85	86	0	255	255
1004	132	133	50	182	182
1005	73	74	9	228	228
1007	92	93	72	255	255
1008	134	135	78	212	212
1009	166	167	11	177	177
1010	157	158	10	167	167
1011	124	125	61	185	185
1012	59	60	36	213	213

Table 3: Histogram properties for PH2 dataset images

Image name	Histogram properties				
	BinCounts	BinEdges	BinLow	BinHigh	BinWidth
IMD002	128	129	0	255	255
IMD003	128	129	0	255	255
IMD004	128	129	0	255	255
IMD006	255	255	0	255	255
IMD008	255	255	0	255	255
IMD009	255	255	0	255	255
IMD010	128	129	0	255	255
IMD014	249	250	6	255	255
IMD015	128	129	0	255	255
IMD016	128	129	0	255	255

Here, Tabs. 1–3 show that the BinLow and BinHigh values are varied mostly except for other histogram properties. In this research, the BinLow and BinHigh values are selected for making the decision. Select the BinLow and BinHigh values for decision. After comparing the histograms of high and low contrast images, we concluded that image contrast variation depends on BinLow and BinHigh values. These two properties are selected for the histogram decision formula. Tab. 4, presents the images with the BinLow and BinHigh values for high and low contrast dermoscopic images. In this research, the BinLow and BinHigh values are selected for calculating the lower (L) and upper (U) limits of the dataset's images.

Table 4: BinLow and BinHigh Values Images

BinLow Value =0 BinHigh Value =210	BinLow Value=30 BinHigh Value=255	BinLow Value =0 BinHigh Value =255	BinLow Value =5 BinHigh Value =255
BinLow Value =38 BinHigh Value =194	BinLow Value =55 BinHigh Value =180	BinLow Value =18 BinHigh Value =220	BinLow Value =15 BinHigh Value =234

Now, the lower limit and upper limit is calculated for all images presents in three datasets. The low and high interval is estimated for analyzing the threshold value of the lower and upper limit. Linear transformation mapping is utilized to calculate the low and high interval. In Fig. 3, the low interval of an image is shown by the histogram. Here, the low interval of an image is between 0 to 60. The design formula computes the low limit for all the images present in datasets explained in step 4.

Like the lower (L) limit, the upper (U) limit is calculated with the help of the BinHigh value of an image in the histogram. The estimated high interval for the upper limit is highlighted in Fig. 3. For this image, the high interval is between 200 to 255. The upper (U) limit is calculated for all dataset images. The upper limit (U) formulation is explained in step 4.

3.2.3 Calculate the Lower (L) and Upper (U) Limits of an Image

The BinLow property of the histogram is used for calculating the lower limit, and the BinHigh property is utilized for upper limit calculation. The lower limit for all the images is calculated by summarizing the average and standard derivation of BinLow values. The average (Avg) denoted as $BinLow_{Avg}$, is calculated by adding all BinLow values and then divided by the total number of images (n), as shown in Eq. (1).

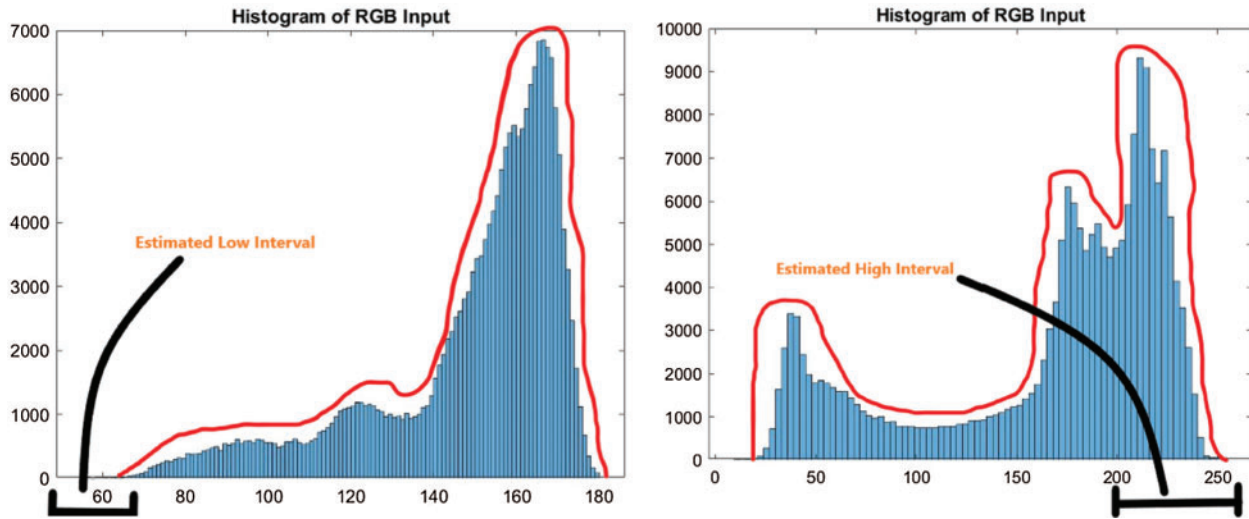$$BinLow_{Avg} = \frac{1}{n} \sum_{i=1}^{n} BinLow_i \tag{1}$$

Figure 3: Estimated low interval of the image histogram

The standard derivation (SD) for all the BinLow values of n number of images is calculated as explained in Eq. (2). Here, n is the number of images of datasets.

$$\text{BinLow}_{\text{SD}} = \sqrt{\frac{\sum (\text{BinLow}_i - \mu)^2}{n}} \tag{2}$$

Then summation of these values is computed from the arithmetic mean, and standard derivation is taken for lower limit (L).

$$L = \text{BinLow}_{\text{Avg}} + \text{BinLow}_{\text{SD}} \tag{3}$$

Same as lower limit calculation, the upper limit is calculated by taking the average and standard derivation of BinHigh values. The average of BinHigh values is computed as seen in Eq. (4).

$$\text{BinHigh}_{\text{Avg}} = \frac{1}{n} \sum_{i=1}^{n} \text{BinHigh}_i \tag{4}$$

The standard derivation (SD) for all the BinHigh values of n number of images is calculated as explained in Eq. (5).

$$\text{BinHigh}_{\text{SD}} = \sqrt{\frac{\sum (\text{BinHigh}_i - \mu)^2}{n}} \tag{5}$$

For upper (U) limit as defined in Eq. (6) of images is calculated by taking the negation after getting the average denoted as $\text{BinHigh}_{\text{Avg}}$ and standard derivation represented as $\text{BinHigh}_{\text{SD}}$ of BinHigh values. The only difference in the upper limit calculation is that the average value's negation and standard derivation are performed.

$$U = \text{BinHigh}_{\text{Avg}} - \text{BinHigh}_{\text{SD}} \tag{6}$$

The contrast evaluation process starts with the computation of histogram decision denoted as HD that measures the lower (L) and Upper (U) limits of images. The HD is formulated as explain in Eq. (7).

$$HD = \sum_{i=0}^{n} f\,(image_i) = \begin{cases} high\ contrast & if\ BinLow < L \vee BinHigh > U \\ low\ contrast & if\ BinLow > L \vee BinHigh < U \end{cases} \qquad (7)$$

where $image_i$ is the number of images present in the dataset, the calculated values for the lower (L) and upper (U) limits are applied in the HD formulation. The lower (L) and upper (U) limits are compared with BinLow and BinHigh values of the input image. After the experiments, the analysis shows that if the BinLow value of an image is less than the lower (L) limit or the BinHigh value of this image is greater than the upper (U) limit. Then it is considering a high contrast image. In the low contrast case, if the BinLow value of an image is greater than the lower (L) limit or the BinHigh value is less than the upper (U) limit, then it is a low contrast image. For dataset ISIC 2017, the high contrast images had the BinLow value less than 34.1 (L), and the Binhigh value is greater than 201.9 (U). While the low contrast images had the BinLow value is greater than 34.1 (L), and the BinHigh value is below 201.9 (U). For dataset ISBI 2016, the high contrast images had the BinLow value less than 30.1 (L), and the BinHigh value is greater than 191.9 (U). The low contrast images had the BinLow value greater than 30.1 (L), and the BinHigh value is below 191.9 (U). For dataset PH2, the high contrast images had the BinLow value less than 6.2 (L), and the BinHigh Value is greater than 246.4 (U). Although the BinLow value is greater than 6.2 (L) for the low contrast images, the BinHigh value is less than 246.4 (U). The proposed histogram decision formulation is based on the BinLow and BinHigh values of an image histogram. After this decision, the skin images are categorized into two contrast variations: (i) high and (ii) low, as described in Tab. 5.

Table 5: Image categorization of dataset ISCI 2017, ISBI 2016, and PH2

ISIC 2017		ISBI 2016		PH2	
Image name	Status	Image name	Status	Image name	Status
ISIC_0000000	High	1002	Low	IMD002	High
ISIC_0000001	High	1003	High	IMD003	High
ISIC_0000002	High	1004	Low	IMD004	High
ISIC_0000004	High	1005	High	IMD006	High
ISIC_0000006	High	1007	Low	IMD008	High
ISIC_0000007	High	1008	Low	IMD009	High
ISIC_0000008	High	1009	Low	IMD010	High
ISIC_0000009	High	1010	Low	IMD014	High
ISIC_0000010	High	1011	Low	IMD015	High
ISIC_0000011	High	1012	High	IMD016	High

Now, the images of the three datasets are separated into two types of images. After making the histogram-based decision, the contrast enhancement or stretching technique is applied only on the low contrast images.

3.3 Pre-Processing for Low Contrast Images

The acquisition of images from three datasets does not suitable for the segmentation process. So, before the lesion segmentation phase, it is necessary to pre-process the input image. The pre-processing phase is an essential step to obtain high accuracy in the next phases. The pre-processing steps for low contrast images are explained in detail. Image contrast stretching enhancement technique is performed on dermoscopic images to improve the contrast of low-quality images. In this technique, the normalization of the image is changed by defining the range for intensity value. From the histogram, change or stretch out the old gray values or levels to the new gray values using the piecewise linear stretching function. The mapping function is used to stretch the values and then create the histogram for the new image. In contrast, stretching technique, the minimum and maximum values that define the image's intensity are increased. After that, the mapping function is implemented to map the histogram's values then the contrast of the image is enhanced. The difference between the low contrast image and contrast stretch image is easily visualized. Now, the skin lesion in the low contrast image is more apparent. Moreover, the similarity problem between background and lesion is also resolved after contrast stretching.

4 Experimental Results

Some fundamental prerequisites are needed to lead this research. Keeping in mind the end goal to diminish processing time, a Graphical Processing Unit (GPU) with higher handling power and more significant memory is best. Be that as it may, constraints to limited hardware access could limit CPU utilization alone. The minimum requirements are the 2.7 GHz processor and 10 GB Random Access Memory (RAM). Moreover, for the software side, this research utilizes MATLAB version R2019a. The operating system used is Windows 10. MATLAB is picked because it employs high-level language and helps programming libraries for numerical calculation, graphics, and programming.

This section encompasses two subsections where Section 4.1 analyses the histogram decision method's performance while Section 4.2 evaluates the performance of image contrast categorization.

4.1 Analysis of Histogram Decision Formulation

The histogram lower (L) and upper (U) limits values are utilized for checking the image contrast. The lower (L) and upper (U) limits are compared with the input image's BinLow and BinHigh values. After experimenting with different values, it was observed that if the BinLow value of an image is less than the lower (L) limit or the BinHigh value of this image is greater than the upper (U) limit, then the input image is considered a high contrast image. In the low contrast case, if the BinLow value of an image is greater than the lower (L) limit or the BinHigh value is less than the upper (U) limit then the input image is a low contrast image. These rules are then applied to all three datasets. For the validation of the research method, the statistics performance measures are used in three datasets. After comparing skin lesion datasets, the stats of high and low contrast images are shown in Tab. 6. The quantity of high contrast images is more than low contrast images. The ISCI 2017 dataset has 900 total images containing 698 high contrast images and 202 low contrast images. The dataset ISBI 2016 includes 910 number of images in which 608 images are high contrast, and 302 number of images are low contrast. Moreover, after the histogram decision, the dataset PH2 contains the 15 number of low contrast images based on

observation. So, the contrast enhancement technique will only be applied to low contrast images. The detailed results of the experiments are discussed below:

Table 6: Comparison of ISIC 2017 and PH2 on high and low contrast images

Phase 1: skin lesion datasets categorization	Dataset		
	ISIC 2017	ISBI 2016	PH2
High contrast images	698	608	185
Low contrast images	202	302	15
Total images	900	910	200

4.2 Performance Measures Results

Two performance measures, the processing time and efficiency, are calculated to evaluate the statistical HD method. The MATLAB time function technique is utilized to measure the processing time of an image in per-processing. This function returns the total time by taking two timestamps, one at the starting and one at the processing image's ending. The processing time is given for, before the HD method and after the HD method as represented in Tab. 7.

Table 7: Processing time before and after statistical histogram decision method

Dataset	Processing time	
	Before the statistical HD method applied	After the statistical HD method applied
ISIC 2017	One image pre-processing time $= 10$ s 900 images pre-processing time $= 900*10$ s $= 9,000$ s $= 2.5$ h ≈ 3 h	One image pre-processing time $= 10$ s 202 image pre-processing time $= 202*10$ s $= 2,020$ s $= 1/2$ h
ISBI 2016	One image pre-processing time $= 10$ s 910 images pre-processing time $= 910*10$ s $= 91,00$ s $= 2.5$ h	One image pre-processing time $= 10$ s 302 image pre-processing time $= 302*10$ s $= 3,020$ s $= 1/2$ h
PH2	One image pre-processing time $= 10$ s 200 images pre-processing time $= 200*10$ s $= 2,000$ s $= 1/2$ h	One image pre-processing time $= 10$ s 15 image pre-processing time $= 15*10$ s $= 150$ s $= 2.5$ min

The graphical representation is also shown in Fig. 4, which shows the processing time improved after the proposed method. The blue color indicates the processing time before the proposed method, and the red color shows the processing time after the proposed method. The low contrast ratio (LCR) is computed from the equation. Here, in the equation, a low contrast image is represented as LCI, and N is the total number of images in the dataset, as seen in Eq. (8). LCI is determined after the categorization of skin lesion datasets using the proposed statistical histogram decision-based method.

$$LCR = \frac{LCI}{N} \times 100 \tag{8}$$

The efficiency of the pre-processing step is also calculated for the evaluation of the proposed method. The efficiency is intended by the equation, as can be seen in Tab. 8. The efficiency in Eq. (9) is denoted as E.
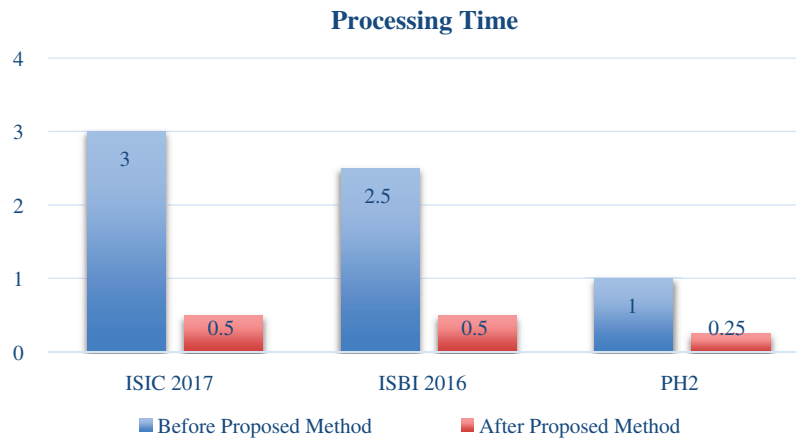
$$E = 100 - LCR \tag{9}$$

Figure 4: Graphical representation of processing time

Table 8: Efficiency evaluation for pre-processing step

Efficiency		
ISIC 2017	ISBI 2016	PH2
$LCR = \frac{202}{900} \times 100 = 22.4\%$	$LCR = \frac{302}{910} \times 100 = 33\%$	$LCR = \frac{15}{200} \times 100 = 7.5\%$
$E = 100 - 22.4 = \mathbf{77}\%$	$E = 100 - 33 = \mathbf{67}\%$	$E = 100 - 7.5 = \mathbf{92.5}\%$

Before the proposed statistical HD method, the processing time and efficiency are not improved. After the proposed method, the efficiency is improved by 77% for ISIC 2017 dataset. The ISBI 2016 dataset displayed 67%. The PH2 dataset showed a 92.5% increase the efficiency.

5 Conclusion

This study proposes a unique perspective to categorize images into low and high by checking the image contrasted through the histogram properties. It is an innovative endeavor to explore image histogram properties to solve pre-processing issues on a vast and diverse dataset. In skin lesion classification, the histogram bin limits values were never utilized before to categorize images into high and low contrast. We have also come up with a statistical formula approach to make the distinction between the images. This has isolated the pre-processing to be done only on a subset of datasets. Due to image categorization, the overall performance is increased by avoiding the extra processing on high contrast images. Through the histogram decision method, the performance of existing skin lesion classification methods is improved. The skin lesion categorization method can be enhanced by incorporating the depth analysis of skin lesion datasets. The histogram's

other properties can be explored like bin width, bin count, and bin height to further research pre-processing on skin lesion datasets.

Conflicts of Interest: The authors declare that they have no interest in reporting regarding the present study.

References

[1] NIH National Cancer Institute, "Cancer facts & figures 2020," *CA. A Cancer Journal for Clinicians*, vol. 31, pp. 1–76, 2020.

[2] F. Afza, M. A. Khan, M. Sharif and A. Rehman, "Microscopic skin laceration segmentation and classification: A framework of statistical normal distribution and optimal feature selection," *Microscopy Research and Technique*, vol. 82, no. 9, pp. 1471–1488, 2019.

[3] R. Javed, M. Shafry, M. Rahim, T. Saba and A. Rehman, "A comparative study of features selection for skin lesion detection from dermoscopic images," *Network Modeling and Analysis in Health Informatics and Bioinformatics*, vol. 1, pp. 1–13, 2020.

[4] A. N. Hoshyar, A. Al-Jumaily and A. N. Hoshyar, "The beneficial techniques in pre-processing step of skin cancer detection system comparing," *Procedia Computer Science*, vol. 42, no. C, pp. 25–31, 2014.

[5] M. Y. Javed, M.A. Khan, M.T. Sharif, T. Saba and A. Rehman, "Multi-model deep neural network based features extraction and optimal selection approach for skin lesion classification," in *2019 Int. Conf. on Computer and Information Sciences (ICCIS)*, Sakaka, Saudi Arabia, IEEE, pp. 1–7, 2019.

[6] A. E. Rad, M. S. M. Rahim, H. Kolivand and M. Amin, "Morphological region-based initial contour algorithm for level set methods in image segmentation," *Multimedia Tools and Applications*, vol. 76, no. 2, pp. 2185–2201, 2017.

[7] A. Rakhmadi, N. Z. S. Othman, A. Bade, M. S. M. Rahim and I. M. Amin, "Connected component labeling using components neighbors-scan labeling approach," *Journal of Computer Science*, vol. 6, pp. 1070–1078, 2010.

[8] R. Javed, Mohd Rahim and S Tanzila, "An improved framework by mapping salient features for skin lesion detection and classification using the optimized hybrid features," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 8, no. 1.6, pp. 95–101, 2019.

[9] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, pp. 60–71, 2019.

[10] T. Saba, "Recent advancement in cancer detection using machine learning: Systematic survey of decades, comparisons and challenges," *Journal of Infection and Public Health*, vol. 13, no. 9, pp. 1274–1289, 2020.

[11] T. Saba, M. A. Khan, A. Rehman and S. L. Marie-Sainte, "Region extraction and classification of skin cancer: A heterogeneous framework of deep CNN features fusion and reduction," *Journal of Medical Systems*, vol. 43, no. 9, pp. 289–311, 2019.

[12] Y. Zhou, C. Shi, B. Lai and G. Jimenez, "Contrast enhancement of medical images using a new version of the world cup optimization algorithm," *Quantitative Imaging in Medicince and Surgery*, vol. 9, no. 9, pp. 1528–1547, 2019.

[13] J. Purushothaman, M. Kamiyama and A. Taguchi, "Color image enhancement based on Hue differential histogram equalization," in *Int. Sym. on Intelligent Signal Processing and Communication Systems*, Phuket, Thailand, pp. 1–5, 2016.

[14] Ş. Öztürk and U. Özkaya, "Skin lesion segmentation with improved convolutional neural network," *Journal of Digital Imaging*, vol. 33, no. 4, pp. 958–970, 2020.

[15] M. M. Rahman, P. Bhattacharya and B. C. Desai, "A multiple expert-based melanoma recognition system for dermoscopic images of pigmented skin lesions," in *8th IEEE Int. Conf. on BioInformatics and BioEngineering*, Athens, Greece, pp. 1–6, 2008.

[16] T. R. Thamizhvani, R. J. Hemalatha, B. Babu, A. Josephin Arockia Dhivya, J. E. Joseph *et al.,* "Identification of skin tumours using statistical and histogram based features," *Journal of Clinical and Diagnostic Research*, vol. 12, no. 9, pp. LC11–LC15, 2018.

[17] M. Rizzi and C. Guaragnella, "Skin lesion segmentation using image bit-plane multilayer approach," *Applied Sciences*, vol. 10, no. 9, pp. 1–15, 2020.

[18] J. Quintana, R. Garcia and L. Neumann, "A novel method for color correction in epiluminescence microscopy," *Computerized Medical Imaging and Graphics*, vol. 35, no. 7–8, pp. 646–652, 2011.

[19] Q. Abbas, I. F. Garcia, M. Emre Celebi, W. Ahmad and Q. Mushtaq, "A perceptually oriented method for contrast enhancement and segmentation of dermoscopy images," *Skin Research and Technology*, vol. 19, no. 1, pp. 490–497, 2013.

[20] R. J. Stanley, R. H. Moss, W. Van Stoecker and C. Aggarwal, "A fuzzy-based histogram analysis technique for skin lesion discrimination in dermatology clinical images," *Computerized Medical Imaging and Graphics*, vol. 23, no. 1, pp. 1–7, 2005.

[21] R. Javed, M. Shafry, M. Rahim, T. Saba and M. Rashid, "Region-based active contour JSEG fusion technique for skin lesion segmentation from dermoscopic images," *Biomedical Research*, vol. 30, no. 6, pp. 1–10, 2019.

[22] P. M. M. Pereira, L. M. N. Tavora, R. Fonseca-Pinto, R. P. Paiva, P. A. A. Assuncao *et al.,* "Image segmentation using gradient-based histogram thresholding for skin lesion delineation," in *Proc. of BIOIMAGING 6th Int. Conf. on Biomedical Engeering Systems and Technologies BIOSTEC*, Rague, Czech Republic, pp. 84–91, 2019.

[23] S. Zehani, A. Toumi, A. Benlamoudi and M. Mimi, "Features extraction using different histograms for texture classification," in *Proc. of ISPA Int. Conf. on Image and Signal Processing and their Applications*, Algeria, 2017.

[24] W. Zhou, S. Gao, L. Zhang and X. Lou, "Histogram of oriented gradients feature extraction from raw bayer pattern images," *IEEE Transactions on Circuits and Systems II Express Briefs*, vol. 67, no. 5, pp. 946–950, 2020.

[25] M. A. Rahman, M. T. Haque, C. Shahnaz, S. A. Fattah, W. P. Zhu *et al.,* "Skin lesions classification based on color plane-histogram-image quality analysis features extracted from digital images," in *IEEE 60th Int. Midwest Symposium on Circuits and Systems (MWSCAS)*, Boston, MA, pp. 1356–1359, 2017.

[26] Y. Faziloglu, R. J. Stanley, R. H. Moss, W. Van Stoecker and R. P. McLean, "Colour histogram analysis for melanoma discrimination in clinical images," *Skin Research and Technology*, vol. 9, no. 2, pp. 147–155, 2003.

[27] J. Premaladha and K. S. Ravichandran, "Novel approaches for diagnosing melanoma skin lesions through supervised and deep learning algorithms," *Journal of Medical Systems*, vol. 40, no. 4, pp. 1–12, 2016.

[28] M. A. Khan, M. I. Sharif, M. Raza, A. Anjum, T. Saba *et al.,* "Skin lesion segmentation and classification: A unified framework of deep neural network features fusion and selection," *Expert Systems*, vol. e12497, pp. 1, 2019.

[29] R. Javed, T. Saba, M. Shafry and M. Rahim, "An intelligent saliency segmentation technique and classification of low contrast skin lesion dermoscopic images based on histogram decision," in *2019 12th Int. Conf. on Developments in eSystems Engineering (DeSE)*, Kazan, Russia, pp. 164–169, 2020.

[30] T. Saba, K. Haseeb, I. Ahmed and A. Rehman, "Secure and energy-efficient framework using Internet of Medical Things for e-healthcare," *Journal of Infection and Public Health*, vol. 13, no. 10, pp. 1567–1575, 2020.