Quantitative Measures for Medical Fundus and Mammography Images Enhancement

Monserrate Intriago-Pazmiño[1,2]*, Julio Ibarra-Fiallo[3], Adán Guzmán-Castillo[2], Raúl Alonso-Calvo[1], José Crespo[1]

[1] Biomedical Informatics Group, Universidad Politécnica de Madrid, Madrid (Spain)
[2] Departamento de Informática y Ciencias de la Computación, Escuela Politécnica Nacional, Quito (Ecuador)
[3] Colegio de Ciencias e Ingeniería, Universidad San Francisco de Quito, Cumbayá (Ecuador)

Received 25 June 2021 | Accepted 23 November 2022 | Early Access 19 December 2022

Abstract

Enhancing the visibility of medical images is part of the initial or preprocessing phase within a computer vision system. This image preparation is essential for subsequent system tasks such as segmentation or classification. Therefore, quantitative validation of medical image preprocessing is crucial. In this work, four metrics are studied: Contrast Improvement Index (CII), Enhancement Measurement Estimation (EME), Entropy EME (EMEE), and Entropy. The objective is to find the best parameters for each metric. The study is performed on five medical image datasets, three retinal fundus sets (DRIVE, ROPFI, HRF-POORQ), and two mammography image sets (MIAS, DDSM). Metrics are calculated using a binary mask image to discard the background. Using the fundus and mask datasets, the best results were obtained with the EMEE and EMEE metrics, which achieved mean improvements of up to 186% and 75%, respectively. For mammography datasets and using masks of the region of interest, the two metrics with the highest percentage improvement were CII and EMEE, which obtained means of up to 396% and 129%, respectively. Based on the experimental results provided, we can conclude that EMEE, EME, and CII metrics can achieve better enhancement assessment in this type of medical imaging.

Keywords

Contrast Improvement Metrics, Contrast Quantitative Measures, Fundus Images, Mammography Images, Medical Image Enhancement.

DOI: 10.9781/ijimai.2022.12.002

I. Introduction

MEDICAL imaging is of great importance in helping specialists to diagnose many diseases. In principle, the specialist analyzes the image, sometimes with the aid of a computer-assisted diagnosis (CAD) system. These CAD systems belong to the field of computer or artificial vision. Artificial vision systems (AVS) for image analysis usually have the phases of preprocessing, segmentation, postprocessing, and feature computation or defect classification [1]. The preprocessing step produces an image of better quality or in a suitable condition for computational analysis in the following phases. The research of this paper focuses on the preprocessing phase. The objective is to study quantitative metrics that measure how much a preprocessed medical image has been improved. The behavior of four metrics on two types of medical images is studied.

The images are enhanced through several filters that modify their contrast and brightness to distinguish the parts of interest. Usually, a qualitative visual analysis is carried out in the preprocessing step, while the validation is focused on later phases. If the outcome is not as expected, the researchers may be forced to go back to the initial phase and try other preprocessing filters.

* Corresponding author.
E-mail address: monserrate.intriago@epn.edu.ec

Concerning the impact of preprocessing tasks on a computer vision system performance, the Master's Thesis [2] examined the effect of preprocessing algorithms on the performance of Convolutional Neural Networks (CNNs) including transfer learning to detecting pneumonia and classifying cats or dogs. This research was conducted using the original images and five enhancement algorithms: the contrast limited adaptive histogram equalization (CLAHE), the successive means of the quantization transform (SMQT), the adaptive gamma correction, the wavelet transform, and the Laplace operator. The chest X-ray and pets' datasets were acquired from Kaggle Challenge. The results reported that LeNet5 CNN performance was improved with some enhancement algorithms, but transfer learning performance slightly decreased with pre-trained VGG16 CNN for pet images.

In contrast, the work [3] obtained better performance when its transfer learning model was tested with enhanced images. The authors studied the impact of enhancing chest X-ray images for COVID-19 detection by applying transfer learning. This study used the large X-ray dataset COVQU, which contains 18,479 CXR images where 8,851 are normal, 6,012 are non-COVID lungs, and 3,616 are COVID-19 CXR images. Additionally, COVQU includes the ground truth lung masks. The study involves classifying each image as normal, lung opacity, or COVID-19 with and without image enhancement. Five image enhancement procedures were used: histogram equalization (HE), contrast limited adaptive histogram equalization (CLAHE), image complement, gamma correction, and balance contrast enhancement technique (BCET). Transfer learning was carried out from six pretrained Convolutional

Please cite this article in press as:
M. Intriago-Pazmiño, J. Ibarra-Fiallo, A. Guzmán-Castillo, R. Alonso-Calvo, J. Crespo. Quantitative Measures for Medical Fundus and Mammography Images Enhancement, International Journal of Interactive Multimedia and Artificial Intelligence, (2022), http://dx.doi.org/10.9781/ijimai.2022.12.002

Neural Networks (CNNs): ResNet18, ResNet50, ResNet101, InceptionV3, DenseNet201, and ChexNet, and the last eleven layers were re-trained. This study achieved seventy-two experiment settings (six pre-retrained CNNs with two datasets, and each dataset has been tested with no enhancement and with five different enhancing methods). The result showed that the image enhancement preprocessing improved the classification performance. Original images (without enhancement) were classified using InceptionV3 obtained 93.46% on accuracy average, and the best experiment setting using gamma correction and ChexNet reached 96.29% on accuracy average.

Other studies also reported higher performance after including some enhancement tasks. A pedestrian detection model based on the YOLOv3 convolutional neural network architecture that analyzes outcomes with and without preprocessing phase is presented in [4]. The preprocessing contrast enhancement is achieved using the Retinex method. The entire proposed model obtained 90% and 94% accuracy, without and with preprocessing, respectively.

In [5], a convolutional neural network (CNN) proposal for recognizing six basic emotions is implemented using several preprocessing methods. The main intention of this study is to investigate how preprocessing practices affect CNN performance. Face detection using a single pre-processing phase achieved a significant result with 86.08 % accuracy. However, combining some techniques increased the performance of CNN and achieved 97.06% accuracy.

The importance of the preprocessing phase in a computer vision system for detecting melanoma has been studied in [6]. Authors recall that preprocessing is the first and fundamental step for improving image quality. In this AVS, preprocessing is designed for removing noise and irrelevant portions against the background of skin photographs. This study applied different pre-processing methods utilized on skin cancer photos. These studies experimentally validated that a suitable preprocessing could increase accuracy.

Most of the research mentioned so far has experimented with the effect of enhancing the images before using them in the learning model of the computer vision system. On the other hand, the image enhancement was not quantified, or at least, it is not reported. To overcome this non-objective scenario, a quantitative image evaluation becomes significant. This research aims to analyze the behavior of quantitative metrics. As a result of this study, this paper recommends appropriate metrics for evaluating the enhancement in each type of image, and studies the values of the relevant algorithms' parameters.

The scientific literature is reviewed in this work, and several metrics used to quantify contrast in medical imaging are studied. Four metrics have been chosen: Enhancement Measure Estimation (EME), Enhancement Measure Estimation by Entropy (EMEE), Contrast Improvement Index (CII), and Entropy. The metrics are applied to two cases of studies: fundus and mammography images, on five data sets. These datasets contain healthy and pathological images. Another point is that some of these datasets include images of poor quality.

The remainder of this article is organized as follows. Section 2 presents a review of related works. Section 3 describes the metrics and their theoretical foundations, the types of images used, and the preprocessing algorithms. Section 4 shows the analysis of the parameters of each metric and its behavior on each dataset. Section 5 provides a discussion of the obtained results. Finally, some conclusions are given.

II. Background

This section presents a review of research works that use metrics to evaluate the performance of the preprocessing filters. Works are presented according to the type of images. First, a cellular medical image work is analyzed; after that, some results corresponding to mammograms are described; finally, studies of retina images of prematurely born children are included.

The work in [7] enhances the contrast of several types of cellular medical images. The enhancement performance is quantified using the CII measure. Cellular images can present complex shapes and textures. Thus, the research concludes that CII is a suitable measure for analyzing the enhancement of these types of images.

A recoloring algorithm, named RGBeat, is presented in [8] in order to assist patients with protanopia and deuteranopia. It is applied to images and text. The proposal uses CIE and RGB color spaces. This method converts a range of values of the hue channel to achieve a better understanding by these types of patients, while preserving the main characteristics. A validation of the modified image is performed. Consistency is addressed by ensuring that all pixels of the same color in an input image will have the same output color after applying the recoloring method. In addition, naturalness is measured by using a quadratic difference in the CIE Lab color space [9]. A small value indicates that the naturalness has been maintained in the recolored image. And the altered contrast measurement is based on a squared Laplacian [10].

The research in [11], which applies deep neural networks for diagnosing congenital heart diseases (CHDs) using echocardiographic ultrasound images, considers the possibility of preprocessing the ultrasound contrast. The authors mention that they are motivated by the work presented in [3], that demonstrated the improved performance of learning methods with contrast-enhanced images.

The contrast of mammography images is enhanced and validated using the CII measure [12]. This research presents a method based on the following filters: Laplacian Gaussian, Contrast Limited Adaptive Histogram Equalization (CLAHE), and morphological filters (openings and closings) to improve the contrast.

A metric based on the high and low-frequency range is proposed in [13]. This metric is used for assessing contrast quality in mammographies. According to the experiments, when the original image is compared with its enhancement, specific conditions of the frequency values are met. The contrast enhanced image has better quality. A private mammographies dataset was used. It contained 179 images, among benign, malignant, and normal mammographs. The study concluded that using multiple filters produces better results and that filter behavior varies between image types.

The machine vision system proposed in [14] aims to recognize the area of distortions in breast images. The distortion classification is mainly performed through an improved pulse-coupled neural network (PCNN). This research utilized the Digital Database for Screening Mammography (DDSM) dataset. In the preprocessing phase, the filters used are top-bottom hat and gamma transformation. Their improvement is validated through the Equivalent Number of Looks (ENL) and Contrast Improvement Index (CII) metrics.

Some measures in [15] are studied to validate contrast enhancement of mammography and tomography pathological images. In this study, ten images for each case are taken. Results are shown in terms of the metrics: EME, EMEE, Logarithmic Michelson Contrast Measure (AME), Logarithmic AME by Entropy (AMEE), Mean Squared Error (MSE), Peak Signal-to-Noise Ratio (PSNR), Absolute Mean Brightness Error (AMBE), Discrete Gray Level Energy (GLE), Relative Entropy (RE), Second-Order Entropy (SOE), Edge Content (EC). The authors classified EME and EMEE as Complex Measures of Contrast, which are convenient metrics for medical images where the background is uniform.

In [16], a method for adjusting the contrast level in a windowed mammogram is proposed. The technique is called GRAIL (Gabor-Relying Adjustment of Image Levels), and it is regulated by a measure of mutual information (MI). MI is the relation between the decomposition

of the original 12-bit inputs and its screen-displayed 8-bit version. The mutual information metric between the original instance and its Gabor-filtered derivations is applied to X-ray images [17], and the results show a better performance in terms of subjective interpretation.

Regarding retinal studies, a new CLAHE version method is presented in [18], and its improvement is evaluated using the Entropy measure, tested on the public DRIVE and STARE fundus image datasets. Those datasets contain adult fundus images of normal and pathological cases.

A private fundus images dataset is used in the research reported in [19]. Authors apply a guided filter for the preprocessing phase of fundus images of children. Healthy and pathological images are used. This filter is assessed through EME, Entropy, Standard Deviation, and Spatial Frequency.

A second work that uses a private set of children's fundus images is [20]. Contrast is enhanced using adaptive filters based on the features of each image. The parameters for each filter are computed employing an artificial neural network. The enhancement is validated using CII and qualitative analysis.

Based on the related works that could be identified at the time of performing this work, the following section details the metrics that are tested, and the datasets of the two case studies that are the object of this research.

III. Materials and Methods

This section describes the metrics, datasets, and preprocessing filters used in this work, which focuses on fundus and mammography. Different algorithms are used to preprocess these types of images to highlight regions of interest. Preprocessing filters reported positively in the literature were selected. Those enhanced images are then used in the subsequent phases of artificial vision systems.

Metrics were implemented using Matlab 2020a. Only the entropy metric is available as a function in Matlab, and the rest of the metrics were coded.

A. Metrics

The metrics described below are applied to two-dimensional images represented as matrixes. The measurement for an entire image is based on partial calculations by blocks without overlapping. Therefore, the first step for computing them is to define a block size and divide the image horizontally and vertically into as many blocks as possible (see Fig. 1). Then, a partial measurement is calculated in each block, and the average of these measurements is considered the value of the metric, taking into account the particularities of each metric.

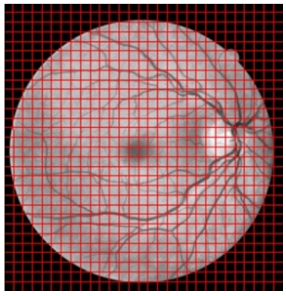


Fig. 1. Example of a fundus medical image divided into square blocks of $L \times L$ size.

1. Enhancement Measure Estimation (EME)

EME is a quantitative measure of contrast enhancement of two-dimensional or grayscale images [21]. The enhancement measure is a modification of Weber's law and Fechner's law. Weber's law establishes that the visual perception of contrast is independent of luminance and low spatial frequency and determines that the perceived change in intensity is proportional to the initial one. On the other hand, Fechner's law states that perception and stimulation are linked logarithmically (i.e., the visually perceived intensity value is proportional to the logarithm of the actual intensity).

For the calculation of the EME metric, the two-dimensional discrete image of $M \times N$ size is divided into small blocks. M and N are the width and the height of the image in pixel number, respectively; the size of the square block is $L \times L$, for $L = 3, 4, ..., n$. Then, the minimum and maximum intensity of each block are found, the contribution of each block is calculated as a natural logarithm function. Finally, the EME value for the whole image is equal to the average. EME is mathematically expressed in (1), where k1 = M/L, is the number of horizontal blocks; k2 = N/L, it is the number of vertical blocks; $I_{max}^{l,m}$ y $I_{min}^{l,m}$ are the maximum and minimum pixel intensity values in a block (m, l), respectively. The size of the block L affects the EME value.

$$EME = \frac{1}{k_1 k_2} \sum_{m=1}^{k_1} \sum_{l=1}^{k_2} 20 \ln\left(\frac{I_{max}^{l,m}}{I_{min}^{l,m}}\right) \quad (1)$$

Note the mathematical equivalence in (2), where $I^{l,m} \in 1, 2, 3, ..., 256$ corresponds to the intensity values shifted by 1 to avoid the indeterminacy of ln(0). So that, the computation of the logarithm is possible, ln $(I^{l,m}) \in \{\ln(1), \ln(2), ..., \ln(256)\}$. That is, ln $(I^{l,m}) \in \{0, 0.6931, ..., 5.5452)\}$. Therefore, the difference between the maximum value and the minimum value remains controlled. Additionally, the expression uses the constant 20 to amplify the difference and provide a higher significance value for EME.

$$ln\left(\frac{I_{max}^{l,m}}{I_{min}^{l,m}}\right) = ln\left(I_{max}^{l,m}\right) - l\,n\left(I_{min}^{l,m}\right) \quad (2)$$

2. Enhancement Measure Estimation by Entropy (EMEE)

This metric is the entropy measure that relates the contrast for each block, scaled by a parameter (α), and averaged over the entire image [21]. The value of α, for $0 < \alpha < 1$, is proportional to the emphasis of the entropy. The variable α helps to manage more randomness. The calculation formula is shown in (3). The impact of the block size (L) and the entropy emphasis (α) over EMEE calculation will be discussed in the Results section.

$$EMEE = \frac{1}{k_1 k_2} \sum_{m=1}^{k_1} \sum_{l=1}^{k_2} \alpha \left(\frac{I_{max}^{l,m}}{I_{min}^{l,m}}\right)^{\alpha} ln\left(\frac{I_{max}^{l,m}}{I_{min}^{l,m}}\right) \quad (3)$$

The factor $\alpha\left(\frac{I_{max}^{l,m}}{I_{min}^{l,m}}\right)^{\alpha}$ highlights the logarithmic amplitude of $I^{l,m}$, increasing the large and compressing minor ones (See Fig. 2). This measure could be more suitable for images that have greater visible variability.

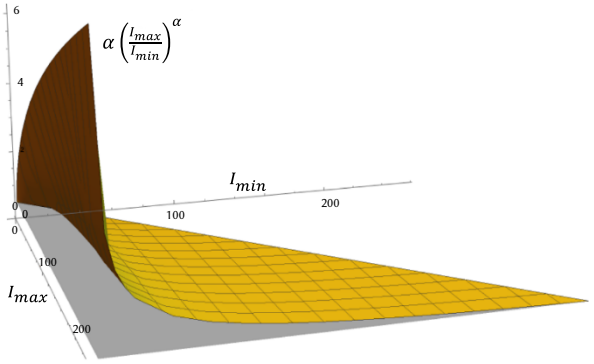


Fig. 2. Weighting factor $\alpha\left(\frac{I_{max}^{l,m}}{I_{min}^{l,m}}\right)^{\alpha}$, $\alpha = 1/2$, $I_{min} \leq I_{max}$.

3. Entropy

Entropy is a statistical measure of randomness. It could characterize textures from an image or photograph [15], [22]. The texture within an image is an element that holds qualities perceived through sight and touch. Texture depicts aspects of the surfaces of photographed objects or subjects. The entropy is calculated on a grayscale image from its histogram. The entropy formula is defined in (4) and (5).

$$Entropy = -\sum_{j=0}^{255} E(I_j) \tag{4}$$

$$E(I_j) = \begin{cases} 0 & if\ P(I_j) = 0 \\ P(I_j)\log_2 P(I_j) & if\ P(I_j) > 0 \end{cases} \tag{5}$$

In the equation (5), $P(I_j) \in [0,1]$ and $P(I_j)$ is the probability of occurrence of the *jth* intensity, according to the image histogram.

4. Contrast Index CI and Contrast Improvement Index (CII)

The Contrast Index (CI) is a measure that establishes the relationship between the background and the foreground of an image. This ratio is calculated by blocks, and the average among these blocks will be the CI measure of the entire image [4]. A high CI value means a more significant difference between background and foreground, and therefore the image presents a good CI. The contrast measure of an image is expressed in (6).

$$CI = \frac{1}{k_1 k_2}\sum_{m=1}^{k_1}\sum_{l=1}^{k_2}\left(\frac{I_{max}^{l,m} - I_{min}^{l,m}}{I_{max}^{l,m} + I_{min}^{l,m}}\right) \tag{6}$$

To establish the improvement of an image, CII is defined in the equation (7). CII is the ratio between the contrast index of the resulting image and the contrast index of the original image. If this ratio is greater than one, the contrast is considered to have been improved.

$$CII = \frac{CI_{processed}}{CI_{original}} \tag{7}$$

B. Fundus Datasets

Fundus images are obtained by examining the retina during pathology inspections. The retinal datasets used in this work contain images of adults and children. Adult datasets are predominant. Although there is a very active scientific community applying machine vision to diagnose child retinal pathologies, there is a lack of public data. Both adults and children may have eye diseases. Some retina diseases are Retinopathy of Prematurity (ROP), Retinopathy of Hypertension, Diabetic Retinopathy, Glaucoma, etc.

1. ROPFI Dataset

Retinopathy of Prematurity Fundus Images (ROPFI) is a set of children's images with Retinopathy of Prematurity [20]. This set entails 64 images captured with a RetCam Shuttle camera (Clarity Medical Systems, Inc.). The average size of an image is 640 x 480 pixels.

2. DRIVE Dataset

Digital Retinal Images for Vessel Extraction (DRIVE) is a public database of adult fundus images [23]. It comprises 40 images whose dimension is 565 x 584 pixels on average. DRIVE is widely used to validate vascular network segmentation algorithms.

3. HRF POOR-QUALITY Dataset

This database contains images of adult patients, and it belongs to a collaborative research group to support comparative studies about automatic segmentation algorithms [24]. It is a set of eighteen images that are of lower quality than others of the same project. The average dimension is 640 x 460 pixels.

4. Masks Acquisition

ROPFI dataset provides the masks that were obtained automatically [20]. Additionally, DRIVE dataset also contains its corresponding masks [23]. While the HRF POOR-QUALITY dataset does not include masks, they have been computed performing the automatic procedure in [20]. First, an Otsu binarization was performed on the green channel. Then a convolution was applied to find borders, and finally the maximum contour was selected to set the binary mask.

C. Mammogram Datasets

1. MIAS Dataset

The Mammographic Image Analysis Society database (MIAS) is a set of 100 labeled and annotated images [25]. The database is suitable for performing and understanding mammograms' technical and visual analysis for research purposes, such as anomalies detection algorithms and other technological derivations that allow computerized assistance to medical specialists. This mammography database is in PGM format, and it contains one hundred images. The average size of the images is 475 x 933 pixels.

2. DDSM Dataset

The Digital Database for Screening Mammography (DDSM) has been published by the University of Florida [26]. DDSM dataset contains 31 mammograms. This dataset is a resource for the mammographic image analysis research community. The main objective of the database is to facilitate research for the development of computer algorithms to assist in the detection of mammography anomalies. It includes the diagnosis and development of assistance aids through software for medical specialists.

3. Automatic Mask Creation

MIAS and DDSM mammogram datasets do not provide masks. Hence, the automatic procedure based on [20] was applied.

D. Filters for Processing Fundus Images

In both infant and adult fundus images, it may not be possible to distinguish attributes of the retina, such as the presence of vessels. It could cause a misdiagnosis by a medical specialist. In [20], this problem is indicated: childhood retinal images present difficulty recognizing retinal elements because they are low in contrast and brightness, have small, curved lines, and present noise. Thus, the authors propose to apply some filters in sequence: contrast, brightness, gamma correction, and CLAHE.

A general image processing operator is a transformation that takes an input I image and computes an output image G. If the image is color, for example, in the red, green, and blue (RGB) color space, the image is usually split into its channels. After applying the filter to each channel, the image can be reconstructed into a color image from its modified channels [27].

The brightness and contrast transformation of one channel at a time can be expressed as in (8):

$$G(x) = c \times I(x)[h] + b \tag{8}$$

The c and b parameters represent the contrast and brightness values that will modify the image, respectively; $I(x)$ describes the intensity image in a pixel x; and h denotes the particular red, green, or blue channel.

Then, a reverse gamma correction filter [27] is applied. This filter removes the non-linear schema of the input image that has been acquired with a digital camera and provides a brighter image using a correction value greater than one. The reverse gamma correction filter is mathematically expressed in (9), where γ is the gamma correction parameter. The gamma transformation is applied to each color image channel. The modified channels are combined, and an image in the same color space as the original image I(x) is obtained.

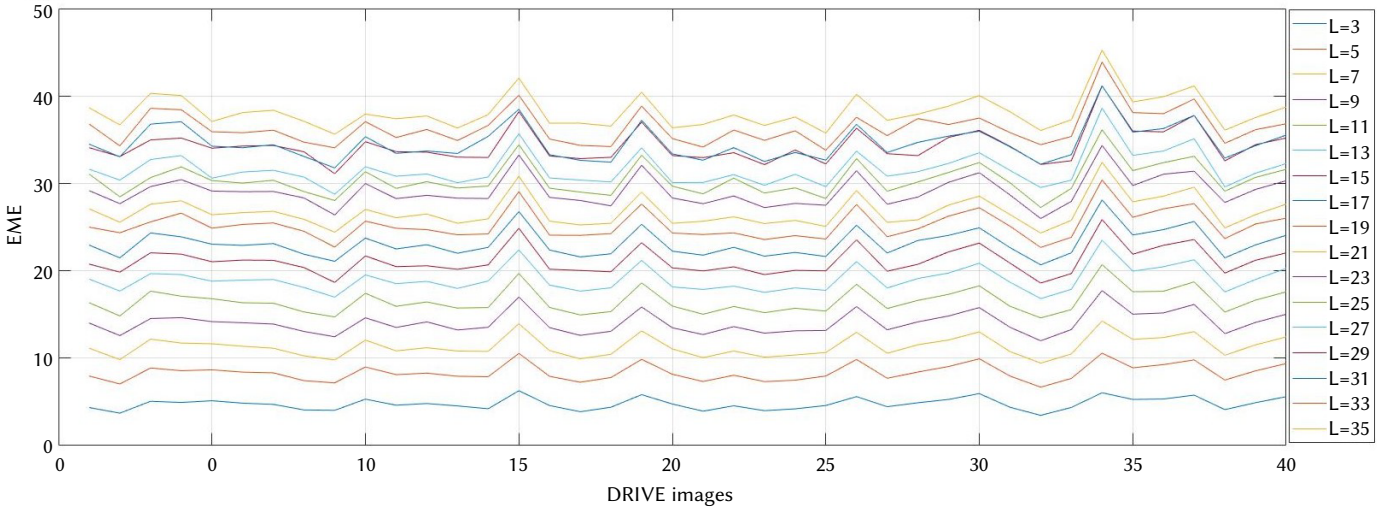


Fig. 3. Computation of EME metric using DRIVE dataset, and varying the block size L. The line color represents a different L value, L = {3, 7, 9, 11, 13, 15, 17, 19, 21, 23, 15, 27, 29, 31, 33, 35} from bottom to top.

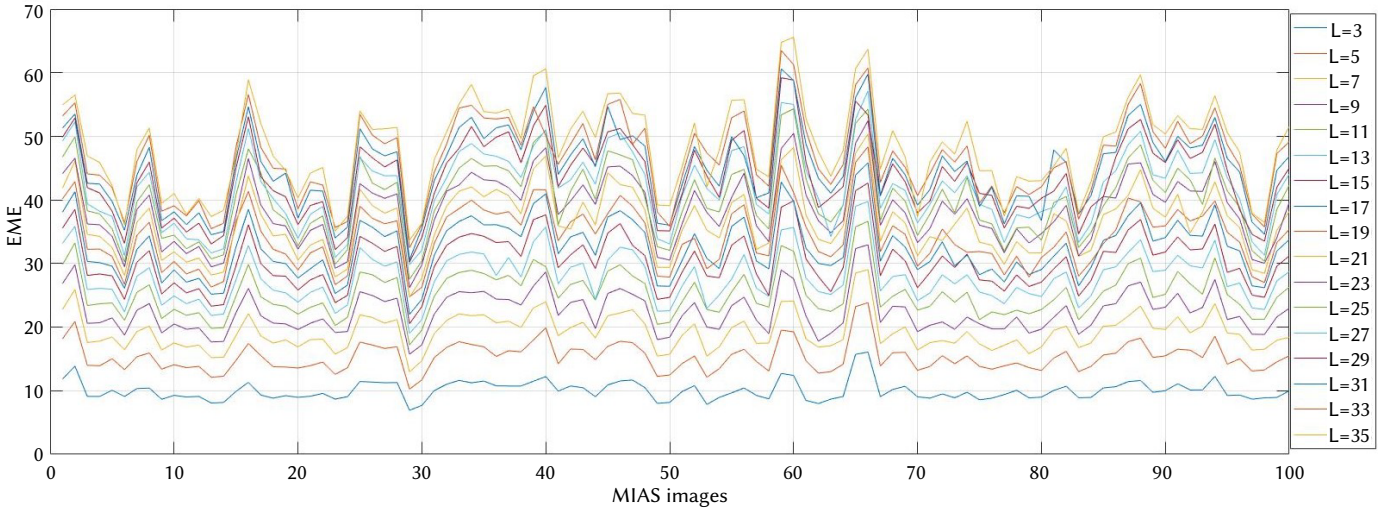


Fig. 4. Computation of EME metric using MIAS dataset, and varying the block size L. The line color represents a different L value, L = {3, 7, 9, 11, 13, 15, 17, 19, 21, 23, 15, 27, 29, 31, 33, 35} from bottom to top.

$$G(x) = I(x)[h]^{\frac{1}{\gamma}} \tag{9}$$

The last filter of the sequence proposed by the authors is the CLAHE [28]. This filter is applied to the L channel of the CIELab color space. The implemented function to process this filter is shown in expression (10), where k and cl represent the variables kernel size and transformation limit, respectively, and L denotes the luminance channel in the corresponding color space.

$$G(x) = CLAHE(I(x)[L], k, cl) \tag{10}$$

E. Filters for Processing Mammogram Images

Mammography images are pre-processed with filters CLAHE y Fast Local Laplacian (FLL) in [29]. The CLAHE filter is used to improve mammography contrast, and FLL is employed to reduce noise and smooth the image. The mathematical representation of CLAHE was given in expression (10). The FLL filter is represented mathematically in (11), where u and v are the coordinates over the x and y axis, respectively, and σ is the standard deviation.

$$G_0 = G_\sigma(u, v) \cdot e^{\frac{-u^2+v^2}{2\sigma^2}} \tag{11}$$

IV. Results

This section begins by analyzing the parameters of each metric and observing how they affect the calculation of the image improvement. Then, the quantitative enhancement of every dataset through each metric is assessed.

A. Analysis of EME Variables

The calculation of EME is based on the size of the block (L). It determines the partial area used to calculate each sum given in (1). EME has been computed for the five datasets shown in the previous section, setting L values from 5 to 19. The analysis of L values is supported by analyzing the plot of the five datasets. As illustrative examples, Fig. 3 and Fig. 4 display the behavior of L in the datasets DRIVE and MIAS, respectively. According to the behavioral analysis of the metric, it has been observed that, with low values of L, the metric reports a low weight as well, which indicates a minor improvement. In contrast, high L values increase the value of the metric. Based on that, and to obtain significant values showing an improvement in the image, a value of L equal to 19 was selected.

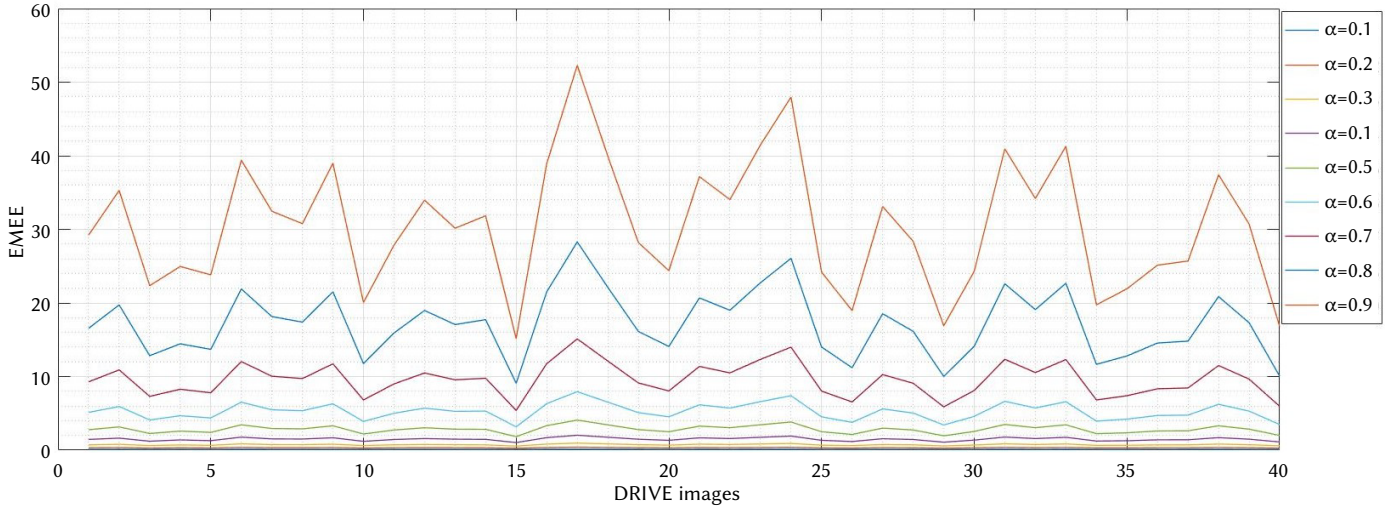


Fig. 5. Computation of EMEE using DRIVE dataset, varying α from 0.1 to 0.9 and keeping block size L=19. The line color represents a different α value, α = {0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.9} from bottom to top.

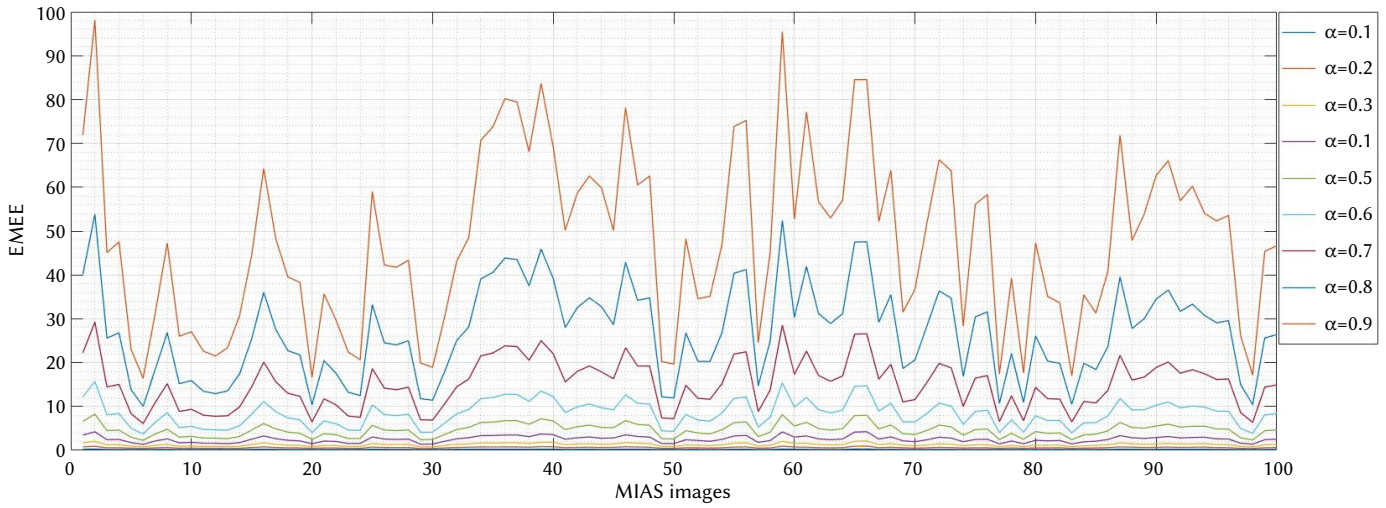


Fig. 6. Computation of EMEE using MIAS dataset, varying α from 0.1 to 0.9 and keeping block size L=19. The line color represents a different α value, α = {0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.9} from bottom to top.

B. Analysis of EMEE Variables

As mentioned above, the EMEE metric has two variables that will influence the calculation: the entropy effect (α) and the block size (L). The behavior of α between 0.1 and 0.9 is analyzed. L is set to the value of 19, as commented in the previous section.

The influence of the variable α for each dataset was analyzed by calculating and plotting EMEE for each dataset. As illustrative examples, Fig. 5 and Fig. 6 show the EMEE metric using both DRIVE of fundus and MIAS of mammograms datasets, respectively. These plots show EMEE as function of α. The EMEE calculation obtained significant values when alpha was set equal to 0.7, 0.8, and 0.9. Taking a middle point alfa has been chosen equal to 0.8. This value will be used for the calculation of the EMME metric.

C. Evaluation of the Improvement on the Fundus Datasets

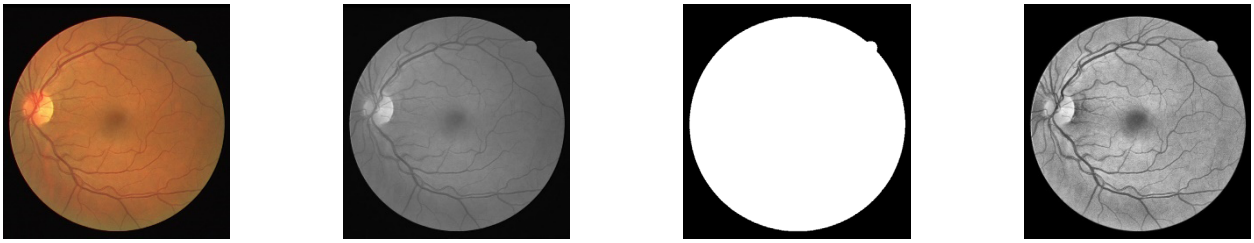
Images from DRIVE, ROPFI, and HRF POOR-QUALITY datasets were improved using the enhancement method proposed in [20], and detailed in Subsection III.D. An important point is to use a mask for delimiting the region of interest (ROI). The mask is a binary image with values 0 and 255, where the pixels with an intensity value equal to 255 constitute the ROI. The discarded part of the image is in black.

Fig. 7, Fig. 8, and Fig. 9 show an image example of DRIVE, ROPFI, and HRF POOR-QUALITY, respectively, that have been preprocessed. Original, original in grayscale, mask, and enhanced images are included. The complete datasets were enhanced with the corresponding algorithm, and subsequently images were evaluated with each metric.

Considering that (6) represents CII as the ratio between the improved image and the original image. Similarly, the ratio of the rest of the measures has been computed. These improvement rates on average for the DRIVE, ROPFI, and HRF POOR-QUALITY datasets are shown in Table I.

TABLE I. Performance of the Fundus Images Enhancement. The Parameters Used Were $L = 19$, and $\alpha = 0.8$. Average of Enhancement (AE) Rate. Average of Enhancement Percentage (AEP)

Dataset	Performance	CII	EME	EMEE	Entropy
DRIVE	**AE Rate**	1.18	1.55	2.37	1.02
	AEP (%)	18.11	**55.42**	**137.47**	1.95
ROPFI	**AE Rate**	1.03	1.75	1.34	1.09
	AEP (%)	3.59	**75.32**	**34.40**	8.98
HRF POOR-Q	**AE Rate**	1.4195	1.66	2.86	1.02
	AEP (%)	41.95	**66.45**	**186.40**	1.95

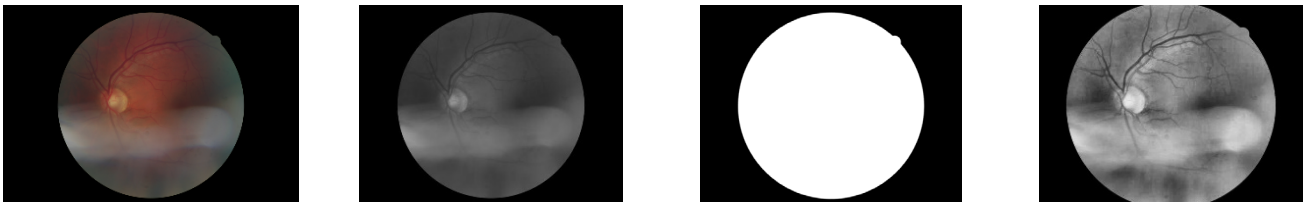


(a) (b) (c) (d)

Fig. 7. First image of the DRIVE dataset. (a) original image, (b) original in grayscale, (c) mask, (d) enhanced image.

(a) (b) (c) (d)

Fig. 8. First image of the ROPFI dataset. (a) original image, (b) original in grayscale, (c) mask, (d) enhanced image.



(a) (b) (c) (d)

Fig. 9. First image of the HRF POOR-QUALITY dataset. (a) original image, (b) original in grayscale, (c) mask, (d) enhanced image.

The average improvement in the DRIVE dataset has been 137%, 55%, 18%, and 2%, applying EMEE, EME, CII, and entropy, respectively. Looking at the report of each metric and considering the averages in Table I, it is possible to conclude that the metrics that achieve the best quantitative distinction are EMEE and EME.

The ROPFI dataset's average improvement percentages reported by EME, EMEE, entropy, and CII have achieved 75%, 64%, 9%, and 4%, respectively. Looking at the report of each metric in Table 1, the metrics that achieve the best quantitative distinction are EME and EMEE.

Concerning the HRF POOR-QUALITY dataset, the improvement percentages on average of 186% and 66% of EME, EMEE, respectively, have achieved the best quantitative distinction.

D. Evaluation of the Improvement on the Mammogram Datasets

As previously mentioned, the MIAS and the DDSM datasets are mammography images. Using the preprocessing method proposed in [27], these sets were improved and commented in Subsection III.E.

Analogously to the evaluation of the fundus images, original and mask images are employed. Each mammography has been preprocessed, and the contrast measurement has been calculated in both the original and preprocessed images. The contrast measurement is only computed for the region of interest using the mask image. The contrast measurement on average for the original and improved images has been calculated, and, finally, the improvement ratio has been obtained. The rate and percentage enhancement on average are stated in Table II.

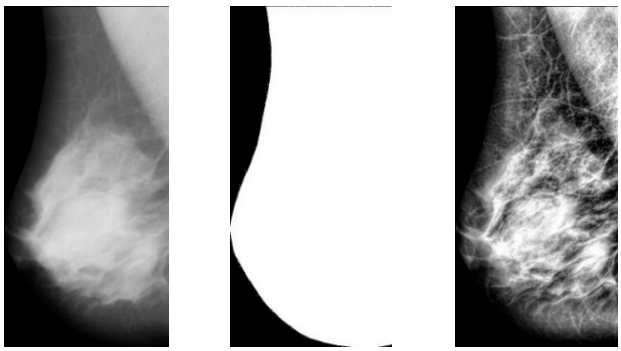
TABLE II. Enhancement Performance of Mammography Datasets. The Parameters Used Were $L = 19$, and $\alpha = 0.8$. Average of Enhancement (AE) Rate. Average of Enhancement Percentage (AEP)

Dataset	Performance	CII	EME	EMEE	Entropy
MIAS	**AE Rate**	4.96	2.70	4.53	1.06
	AEP (%)	**396.55**	170.11	**353.39**	6.05
DDSM	**AE Rate**	2.30	1.57	1.82	1.04
	AEP (%)	**129.92**	56.97	**82.04**	4.16

A pair of improved images of the MIAS and DDSM datasets are presented in Fig. 10 and 11, respectively. These figures incorporate original in color, original in grayscale, mask, and preprocessed, respectively.

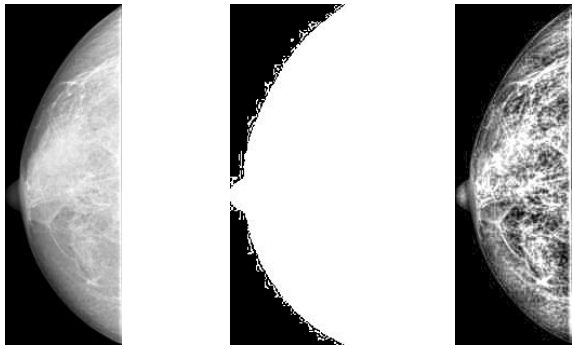
In the case of the MIAS dataset, CII and EMME indicated the highest enhancement values, 396% and 353% on average, respectively. EME also presents a high average value with respect to the Entropy average, 160%, and 2%, respectively.

With respect to the DDSM dataset, CII is the most significant value, which reported 129% on enhancement average. EMEE, EME, and entropy registered 66%, 57%, and 4% on enhancement average, respectively. Thus, EMEE is close to the EME value, and both are quite large with respect to the Entropy.



(a) (b) (c)

Fig. 10. First image of the MIAS dataset. (a) original image, (b) mask, (c) enhanced image.



(a) (b) (c)

Fig. 11. First image of the DDSM dataset. (a) original image, (b) mask, (c) enhanced image.

V. Discussion

In a computer vision system, contrast and brightness enhancement of images is part of the first phase named preprocessing, and it is essential for the following phases. This research deals with fundus images of the retina and mammograms, including healthy and pathological cases. The improvement algorithms could be assessed through experts' criteria and quantitative validation. Expert criteria agreement may require the analysis of several experts and could be subjective.

In the literature review, it has been commented that it is not common in artificial vision systems to report the image improvement quality using quantitative metrics. It seems that scientists rely on visual perception or on trial and error. On the other hand, there have been a few works that have been considered quantitative aspects, as discussed in previous sections.

Subsequent phases of the computer vision system may receive an image with a non-satisfactory enhancement, and brightness and contrast alteration may need to be performed again. Consequently, appropriate quantitative measures are quite valuable.

In this research, medical images of retina fundus and mammograms have been selected to study and quantify contrast measurements. These two types of pathological images are of broad interest to physicians and informatics specialists. Because fundus images usually are colored, and mammograms are grayscale, these datasets permit to evaluate the proposed measures in both colored and grayscale images. Regarding the similarities, it can be observed that both types examine anatomical parts of the human body, and that the background is a large portion of the image (so that binary masks are used to delimit the area of interest). Because of these common characteristics, it was possible to examine both datasets using similar scripts.

Previous works agree that a high metric value represents a better distinction of the parts of interest versus the background [2], [6], [12]. However, the parameter's values were not reported. Due to this, the first effort of this work has been to establish the most suitable parameters of each metric through mathematical analysis and experimentation. Those most suitable parameters have been identified as the block size ($L = 19$), and the entropy emphasis, ($\alpha = 0.8$), as treated in the Results Section.

In our case studies, in mammograms, CII and EMEE reported the highest contrast enhancement rates of up to 396% and 353%, respectively, and EME of up to 170%; and, regarding retinal datasets, EMEE, CII, and EME metrics reported enhancements of up to 186%, 75%, and 41%, respectively. Entropy is the measure with the smallest margin of distinction in both fundus and mammography images. However, there is a high improvement ratio in the case of mammograms compared to fundus images. The improvement percentage of the entropy metric ranges from 2% to 8%. Accordingly, it can be recommended using the EMEE, EME and CII metrics to quantitatively validate the contrast and brightness improvement of medical images.

An analysis of the behavior of the measurement as a function of parameters L and α was carried out. This being so, the chosen parameters allow differentiating better the image improvement. The values of each parameter have been studied and reported precisely, with the intention that researchers who need to use those metrics know the most convenient parameters.

In Section I, it was commented that the image enhancement algorithm influences and improves, in most cases, the performance of the artificial vision system. We have presented how the metrics can achieve more significant improvements in certain types of images. Therefore, for images similar to the cases studied in the paper, the preprocessing and metric values presented could be applied. For other, quite different images, the analysis and guidelines presented in the paper can be adapted to perform the analysis and parameter selection.

The scope of this work has been to study evaluation metrics of image enhancement algorithms. Also, this work has been considered mammography and fundus images. In order to include various types of healthy and pathological images, three different fundus datasets and two different mammography datasets were included.

This research work could be most valuable for researchers that develop computer vision applications, in order to evaluate the quality of their preprocessed images and improve the applicability of their techniques.

Since the amount and variety of datasets have not been extensive, the main future works are to extend this research by evaluating other sets of related medical images, reproduce a complete computer vision method, and report the relation between quantitative enhancement and the computer vision system performance.

VI. Conclusion

The review of related works indicated that the image pre-processing phase affects the results achieved by subsequent steps of an artificial vision system. As reported, the correct preprocessing of the input images accomplished that deep neural network techniques could improve up to 4% their accuracy. Thus, the consideration of this early quantitative assessment of image quality could be incorporated into the design of machine vision systems in the medical imaging field.

The need to quantitatively validate the enhancement of medical images in the first phase (or preprocessing) of a computer vision system was a main motivation of this research work. And, as discussed in the paper, metrics EMEE, EME and CII are valuable for measuring the enhancement of the studied medical images.

To apply these metrics in new datasets, an analysis of the metrics parameters following the approach of this paper is recommended. An important consideration is that the region of interest of images should be satisfactorily delimited.

In future work, it is planned to initiate a collaboration with additional clinical specialists to gather their opinions and suggestions about the preprocessing phase, so that they could be taken into account in future developments.

Acknowledgment

We thank Metrofraternidad Foundation, Quito, Ecuador, for providing the ROPFI dataset.

References

[1] Y. Boutiche, "Fast level set algorithm for extraction and evaluation of weld defects in radiographic images," *Studies in Computational Intelligence*, vol. 672, pp. 51–68, 2017, doi: 10.1007/978-3-319-46245-5_4.

[2] X. Chen, "Image enhancement effect on the performance of convolutional neural networks", M.S. thesis, Faculty of Computing, Blekinge Institute of Technology, Karlskrona, Sweden, 2019, doi:10.1016/j.compbiomed.2021.104319.

[3] T. Rahman, et. al, "Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images", Computers in Biology and Medicine, vol. 132, 2021.

[4] H. Qu, T. Yuan, Z. Sheng, and Y. Zhang, "A Pedestrian Detection Method Based on YOLOv3 Model and Image Enhanced by Retinex," *Proceedings - 2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, CISP-BMEI 2018*, Feb. 2019, doi: 10.1109/CISP-BMEI.2018.8633119.

[5] D. A. Pitaloka, A. Wulandari, T. Basaruddin, and D. Y. Liliana, "Enhancing CNN with Preprocessing Stage in Automatic Emotion Recognition," *Procedia Computer Science*, vol. 116, pp. 523–529, Jan. 2017, doi: 10.1016/J.PROCS.2017.10.038.

[6] E. Vocaturo, E. Zumpano and P. Veltri, "Image pre-processing in computer vision systems for melanoma detection," 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2018, pp. 2117-2124, doi: 10.1109/BIBM.2018.8621507.

[7] J. P. Gu, L. Hua, X. Wu, H. Yang, and Z. T. Zhou, "Color medical image enhancement based on adaptive equalization of intensity numbers matrix histogram, "*International Journal of Automation and Computing*, vol. 12, no. 5, pp. 551–558, 2015, doi: 10.1007/s11633-014-0871-9.

[8] M. M. G. Ribeiro, and A. J. P. Gomes, "RGBeat: A Recoloring Algorithm for Deutan and Protan Dichromats, "*International Journal of Interactive Multimedia and Artificial Intelligence*, In Press, pp. 1–13, 2022, doi: 10.9781/ijimai.2022.01.003.

[9] D. R. Flatla, K. Reinecke, C. Gutwin, and K. Z. Gajos, "SPRWeb: preserving subjective responses to website colour schemes through automatic recolouring," *in Proc. Conf. Human Factors in Computing Systems (SIGCHI'13). ACM*, 2013, pp. 2069–2078, doi: 10.1145/2470654.2481283.

[10] X. Xu, Y. Wang, J. Tang, X. Zhang, and X. Liu, "Robust automatic focus algorithm for low contrast images using a new contrast measure," Sensors, vol. 11, no. 9, pp. 8281–8294, 2011, doi: 10.3390/s110908281.

[11] S. Chen, C. Wang, I. Tai, K. W, Y. Chen, and K. Hsieh, "Modified YOLOv4-DenseNet Algorithm for Detection of Ventricular Septal Defects in Ultrasound Images," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 6, no. 7, pp. 101–108, 2021, doi: 10.9781/ijimai.2021.06.001.

[12] S. Wu, Q. Zhu, Y. Yang, and Y. Xie, "Feature and contrast enhancement of mammographic image based on multiscale analysis and morphology," *2013 IEEE International Conference on Information and Automation, ICIA 2013*, vol. 2013, pp. 521–526, 2013, doi: 10.1109/ICInfA.2013.6720354.

[13] A. Pandey and S. Singh, "New performance metric for quantitative evaluation of enhancement in mammograms," *Proceedings of the 2013 2nd International Conference on Information Management in the Knowledge Economy, IMKE 2013*, pp. 51–56, 2014.

[14] G. Du *et al.*, "A new method for detecting architectural distortion in mammograms by nonsubsampled contourlet transform and improved PCNN," *Applied Sciences (Switzerland)*, vol. 9, no. 22, 2019, doi: 10.3390/app9224916.

[15] S. Gupta and R. Porwal, "Appropriate Contrast Enhancement Measures for Brain and Breast Cancer Images," *International Journal of Biomedical Imaging*, vol. 2016, no. 1, 2016, doi: 10.1155/2016/4710842.

[16] A. Albiol, A. Corbi, and F. Albiol, "Automatic intensity windowing of mammographic images based on a perceptual metric," *Medical Physics*, vol. 44, no. 4, pp. 1369–1378, Apr. 2017, doi: 10.1002/MP.12144.

[17] A. Albiol, A. Corbi, and F. Albiol, "Measuring X-ray image quality using a perceptual metric," *2016 Global Medical Engineering Physics Exchanges/Pan American Health Care Exchanges, GMEPE/PAHCE 2016*, Jul. 2016, doi: 10.1109/GMEPE-PAHCE.2016.7504639.

[18] K. Aurangzeb, S. Aslam, M. Alhussein, R. A. Naqvi, M. Arsalan, and S. I. Haider, "Contrast Enhancement of Fundus Images by Employing Modified PSO for Improving the Performance of Deep Learning Models," *IEEE Access*, vol. 9, 2021, doi: 10.1109/ACCESS.2021.3068477.

[19] K. L. Nisha, S. G., P. S. Sathidevi, P. Mohanachandran, and A. Vinekar, "A computer-aided diagnosis system for plus disease in retinopathy of prematurity with structure adaptive segmentation and vessel based features," *Computerized Medical Imaging and Graphics*, vol. 74, pp. 72–94, 2019, doi: 10.1016/j.compmedimag.2019.04.003.

[20] M. Intriago-Pazmino, J. Ibarra-Fiallo, J. Crespo, and R. Alonso-Calvo, "Enhancing vessel visibility in fundus images to aid the diagnosis of retinopathy of prematurity," *Health Informatics Journal*, pp. 1–15, 2020, doi: 10.1177/1460458220935369.

[21] S. S. Agaian, K. P. Lentz, A. M. Grigoryan, S. S. Agaian, K. P. Lentz, and A. M. Grigoryan, "A New Measure of Image Enhancement," *IASTED International Conference on Signal Processing & Communication*, no. January 2000, pp. 19–22, 2000.

[22] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 4th ed. New York: Pearson, 2018.

[23] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, "Ridge-Based Vessel Segmentation in Color Images of the Retina," *IEEE Transactions on Medical Imaging*, vol. 23, no. 4, pp. 501–509, Apr. 2004, doi: 10.1109/TMI.2004.825627.

[24] A. Budai, R. Bock, A. Maier, J. Hornegger, and G. Michelson, "Robust vessel segmentation in fundus images," *International Journal of Biomedical Imaging*, vol. 2013, 2013, doi: 10.1155/2013/154860.

[25] J. Suckling, J; Parker, J; Dance ,D; Astley, S; Hutt, I; Boggis, C; Ricketts, I; Stamatakis, E; Cerneaz, N; Kok, S; Taylor, P; Betal, D; Savage, "The Mammographic Image Analysis Society Digital Mammogram Database," *Kluwer Academic Publishers*, vol. 13, pp. 375–378, 1994.

[26] K. Bowyer, K., Kopans, D., Kegelmeyer, W. P., Moore, R., Sallam, M., Chang, K., Woods, "HeathEtAl2001_DDSMdataset," in *In Third international workshop on digital mammography*, 1996, p. 27.

[27] R. Szeliski, *Image formation*, vol. 73. 2011. doi: 10.1007/978-3-642-2014

[28] K. Zuiderveld, *Contrast Limited Adaptive Histogram Equalization*. Academic Press, Inc., 1994. doi: 10.1016/b978-0-12-336156-1.50061-6.

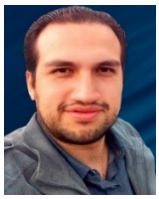
[29] N. Arora and G. Aggarwal, "Mammogram Classification Using Genetic Algorithmbased Feature Selection," *International Journal of Research in Electronics and Computer Engineering a Unit of I2OR*, vol. 6, no. 4, pp. 378–379, 2018.

Monserrate Intriago-Pazmiño

She received the B.S. degree in Computer Science Engineering from National Polytechnic School, Quito, Ecuador, in 2007, and the M.S degree in Computer Science from the Technical University of Madrid, Madrid, Spain, in 2012. She is currently an Associate Professor at the Department of Informatics and Computer Science, at the National Polytechnic School. She is a Ph.D. candidate in Computer Science at Technical University of Madrid. Her research interests include software quality and machine learning.

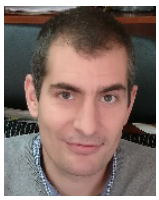
Julio Ibarra-Fiallo

He received the degree of Mathematician from the National Polytechnic School, Ecuador, and the master's degree in Applied Mathematics, from the San Francisco University of Quito, Ecuador. He is currently an Associate Research Professor at the College of Sciences and Engineering at the San Francisco University of Quito. He has participated and directed national research projects. He researches in areas of artificial intelligence, pattern recognition, computer vision, typical testors, combinatorial logic models of variable selection, neural networks, among others.



Adán Guzmán-Castillo

He received the Engineering degree in Computer and Information Systems from the Escuela Politécnica Nacional (EPN), Ecuador, in 2020. He is currently pursuing the M.Sc. degree at the Escuela Politécnica Nacional (EPN). He is also a Research Technician with EPN. His current research interests include studies of medical image segmentation algorithms and studies of interoperability solutions in IoT environments involving multiple devices working with different technologies.



Raúl Alonso-Calvo

He is a PhD in Computer Science from the Universidad Politécnica de Madrid (UPM). He has been a visitor researcher at Universidade de Aveiro DETI-IEETA (Portugal) (2010), and Oxford University (UK) (2014). Since 2010 he is at the Departamento de Lenguajes Sistemas Informáticos e Ingeniería de Software, ETSI Informáticos at UPM, where he is currently an Associate Professor. He has been a member of the Biomedical Informatics Group at UPM from 2001. His research interests are mainly focused on clinical research informatics, biomedical interoperability standards, database integration and preprocessing, biomedical image processing and information retrieval in biomedicine. He has been author and co-author of research papers in several journals. He has participated in EU research projects since 2001 and in recent years he was involved in INTEGRATE: Driving Excellence in Integrative Cancer and EURECA: Enabling information re-use by linking clinical Research and Care.

José Crespo

He received the degrees of Master of Science (School of Electrical and Computer Engineering), Master of Science (School of Management), and the Ph.D. degree (School of Electrical and Computer Engineering) from the Georgia Institute of Technology (Atlanta, USA). He received the degree of "Ingeniero de Telecomunicaciones" from "Universidad Politécnica de Madrid" (Spain). He is a full Professor at "ETS Ingenieros Informáticos", "Universidad Politécnica de Madrid" (Spain), where he belongs to the Biomedical Informatics Group. He has participated in national and international research projects. He has been Head of Department.