



# Article Bio-Inspired Watermarking Method for Authentication of Fundus Images in Computer-Aided Diagnosis of Retinopathy

Ernesto Moya-Albor<sup>1,\*,†</sup>, Sandra L. Gomez-Coronel<sup>2,†</sup>, Jorge Brieva<sup>1,\*,†</sup> and Alberto Lopez-Figueroa<sup>1,†</sup>

- <sup>1</sup> Facultad de Ingeniería, Universidad Panamericana, Augusto Rodin 498, Ciudad de México 03920, Mexico; 0234621@up.edu.mx
- <sup>2</sup> Instituto Politécnico Nacional, UPIITA, Departamento de Ingeniería, Av. IPN No. 2580, Col. La Laguna Ticomán, Ciudad de México 07340, Mexico; sgomezc@ipn.mx
- \* Correspondence: emoya@up.edu.mx (E.M.-A.); jbrieva@up.edu.mx (J.B.); Tel.: +52-55-5482-1600 (ext. 5210) (E.M.-A.)
- <sup>†</sup> These authors contributed equally to this work.

Abstract: Nowadays, medical imaging has become an indispensable tool for the diagnosis of some pathologies and as a health prevention instrument. In addition, medical images are transmitted over all types of computer networks, many of them insecure or susceptible to intervention, making sensitive patient information vulnerable. Thus, image watermarking is a popular approach to embed copyright protection, Electronic Patient Information (EPR), institution information, or other digital image into medical images. However, in the medical field, the watermark must preserve the quality of the image for diagnosis purposes. In addition, the inserted watermark must be robust both to intentional and unintentional attacks, which try to delete or weaken it. This work presents a bio-inspired watermarking algorithm applied to retinal fundus images used in computer-aided retinopathy diagnosis. The proposed system uses the Steered Hermite Transform (SHT), an image model inspired by the Human Vision System (HVS), as a spread spectrum watermarking technique, by leveraging its bio-inspired nature to give imperceptibility to the watermark. In addition, the Singular Value Decomposition (SVD) is used to incorporate the robustness of the watermark against attacks. Moreover, the watermark is embedded into the RGB fundus images through the blood vessel patterns extracted by the SHT and using the luma band of Y'CbCr color model. Also, the watermark was encrypted using the Jigsaw Transform (JST) to incorporate an extra level of security. The proposed approach was tested using the image public dataset MESSIDOR-2, which contains 1748 8-bit color images of different sizes and presenting different Diabetic Retinopathy (DR). Thus, on the one hand, in the experiments we evaluate the proposed bio-inspired watermarking method over the entire MESSIDOR-2 dataset, showing that the embedding process does not affect the quality of the fundus images and the extracted watermark, by obtaining average Peak Signal-to-Noise Ratio (PSNR) values higher to 53 dB for the watermarked images and average PSNR values higher to 32 dB to the extracted watermark for the entire dataset. Also, we tested the method against image processing and geometric attacks successfully extracting the watermarking. A comparison of the proposed method against state-of-the-art was performed, obtaining competitive results. On the other hand, we classified the DR grade of the fundus image dataset using four trained deep learning models (VGG16, ResNet50, InceptionV3, and YOLOv8) to evaluate the inference results using the originals and marked images. Thus, the results show that DR grading remains both in the non-marked and marked images.

**Keywords:** image watermarking; Steered Hermite transform; Jigsaw transform; singular value decomposition; spread spectrum; fundus images; retinopathy; convolutional neural networks; YOLO; information security

MSC: 68U10; 68T05; 92C55



Citation: Moya-Albor, E.; Gomez-Coronel, S.L.; Brieva, J.; Lopez-Figueroa, A. Bio-Inspired Watermarking Method for Authentication of Fundus Images in Computer-Aided Diagnosis of Retinopathy. *Mathematics* 2024, 12, 734. https://doi.org/10.3390/ math12050734

Academic Editors: Cheonshik Kim and Ki-Hyun Jung

Received: 1 February 2024 Revised: 24 February 2024 Accepted: 25 February 2024 Published: 29 February 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). The World Health Organization (WHO) classifies visual impairment into two groups (International Classification of Diseases 11 (2018)): visual impairment of distance and near presentation. For the distance vision disability group, blindness is considered when the person has a visual acuity of less than 3/60, considering a normal visual acuity of 20/20 [1].

Globally, at least 2.2 billion people have near or far vision impairment, whereas in at least 1 billion people the vision impairment could have been prevented or has not yet been treated. In addition, of these 1 billion people, about 3.9 million people have Diabetic Retinopathy (DR) as the main cause of their visual impairment [1].

In general, the term retinopathy is a generic term used to refer to any non-inflammatory disease affecting the retina, being used to group a set of different conditions, each with its characteristics. The most common retinopathies are retinopathy of prematurity, diabetic retinopathy, hypertensive retinopathy, and central serous retinopathy [2].

In Latin America and the Caribbean, there is a high prevalence of diabetes, in 2015 it was estimated that 29.6 million people were living with diabetes. In addition, it is estimated that more than 75% of patients who have had diabetes mellitus for more than 20 years will have some form of diabetic retinopathy, which is responsible for 2.6% of blindness worldwide. On the other hand, it is estimated that after 15 years of diabetes, approximately 2% of patients will become blind and 10% will develop severe visual impairment [3,4]. Particularly, it is estimated that in Mexico there is a 71% of DR incidence in diabetic patients [5].

Currently, some treatments can significantly reduce the risks of blindness and moderate vision loss by more than 90% [3]. In this sense, fluor angiography is one of the techniques that has contributed to the understanding, diagnosis, and treatment of many chorioretinal diseases [6], which consists of the intravenous administration of a dye substance called fluorescein and photographs are taken, using a special camera, of the back of the eye to assess how the dye flows in the arteries, capillaries, and veins of the inner part of the eye, known as fundus images. However, there is the possibility of developing adverse reactions such as urticaria, fever, and chills that correspond to individual susceptibility [6]. On the other hand, retinography is a lower cost non-invasive diagnostic technique that does not use contrast agents in the acquisition process avoiding possible reactions. This technique produces RGB fundus images that can be used to detect some of the most distinctive structures that characterize retinopathies, such as neovascularization, hemorrhages, exudates, and microaneurysms [7].

Fundus imaging is a process in which the 3D structure of the retina is projected onto the 2D plane. The intensity of the image represents the amount of reflected light. The fundus camera consists of a low-power microscope and a camera attached to the top of the microscope. The camera can capture the retinal area at an angle of  $30^{\circ}$  to  $50^{\circ}$  with a magnification of  $2.5 \times (5 \times$  using auxiliary lenses). Color filters, fluorescein, and indocyanine green dyes are used to obtain the fundus image. There are mainly three modalities for fundus photography of the retina [8]:

- 1. Fundus photography (red-free): A wavelength band is used to capture the amount of reflected light.
- 2. RGB fundus photography: The red (R), green (G), and blue (B) wavelength bands are used to capture the amount of reflected light.
- 3. Fluorescein angiography (fluorangiography) and indocyanine: The image is generated from the amount of photons emitted by the fluorescein and indocyanine dyes injected into the patient.

Another area of study in recent years is the vulnerability of patient digital information, hence, copyright protection. For example, steganography and image watermarking have been widely used to hide digital information in a cover image. In the case of watermarking, several methods have been developed with the objective of these methods to insert information into the cover image. Depending on the application, each method uses different tools to achieve its goal and they must consider the requirements to design the algorithm: imperceptibility, robustness, legibility, ambiguity, and security. However, discerning when

a watermarking method is focused on medical images is also very important to preserve the diagnosis after, including the watermark and protecting the watermark. In this case, generally, the watermark includes is information about the patient, i.e., the Electronic Patients Record (EPR), but in some cases, authors use QR codes, logos, etc., as a watermark. So, to achieve these elements, designing an imperceptible watermarking algorithm is usually required. The state-of-the-art indicated that the proposals can be designed in the spatial domain, the transform domain, and in recent years with hybrid techniques. The spatial domain techniques have the advantage of low computational cost, but because pixels are directly modified, the images suffer visible modifications, altering the diagnosis. So it is preferable to design a method in the transform domain or combine different transforms (hybrid method). The most popular transforms employed are the Discrete Fourier Transform (DFT), Discrete Cosine Transform (DCT), and Discrete Wavelet Transform (DWT). In addition, when algorithms are focused on medical images, the state-of-the-art indicates classified methods. According to [9], medical image watermarking methods can be classified as Region Of Interest (ROI)-based watermarking, reversible watermarking, and imperceptible watermarking. In [10], the authors indicated the next classification: ROI, reversible watermarking, and zero watermarking. Finally, the authors of [11] indicated that the watermark algorithms for medical images could be classified as Region Of No Interest (RONI), reversible watermarking, and conventional digital watermarking. As we can see, there are algorithms focused on determining the region of interest to insert the watermark. So, different proposals to watermark digital images consider the classification to define their algorithm as a hybrid because they combine it, i.e., the proposals that combine RONI with a reversible schema sometimes are not successful because trying to recover the watermark fails in the manipulated areas by protecting the integrity of the image.

Regarding the watermarking of medical images, a very important point is not only to evaluate a watermarking method in terms of robustness, security, and imperceptibility but also to evaluate the proposal with real medical images and determine if the real diagnosis changes or not. Different watermark algorithms use retinal fundus images and evaluate if the method modified the diagnosis or not. For example, ref. [9] presents an imperceptible watermark method for medical images and the authors include tests demonstrating that it does not affect computer vision-based automated diagnosis of retinal diseases. On the other hand, Dey et al. [12] reported a watermarking algorithm for fundus images by inserting patient information, and evaluated that the embedding watermark process slightly modifies the blood vessel extraction.

In this work, we propose a hybrid watermark method based on the Steered Hermite Transform (SHT), the Singular Value Decomposition (SVD), and the Jigsaw Transform (JST) applied to a public dataset of RGB fundus images presenting a grade of diabetic retinopathy. Thus, the SHT, a bio-inspired image model, is used as a spatial frequency decomposition tool to embed the watermark, providing its imperceptibility; the JST, a popular image scrambling technique, is used to increase the security; and the SVD generates robust watermarks against attacks. Additionally, we evaluate, through a deep learning classification strategy, if the watermarking algorithm applied to these medical images modifies the diagnosis.

The rest of the document is structured as follows: Section 2 presents a revision of the state-of-the-art works, both those for watermarking applied to medical images, in particular to fundus images and those deep learning-based works used as support to the diagnosis of diabetic retinopathy. Section 3 reports the fundus image dataset and the watermark used to test the watermarking approach and used to classify the DR grade through deep learning models. In addition, this section presents an overview of the proposed method and introduces the Steered Hermite transform, the Jigsaw transform, the SVD, and the basis of the Convolutional Neural Networks (CNN). On the other hand, Section 4 describes the proposed bio-inspired watermarking method, reporting both the insertion and extraction processes, as well as the trained deep learning models used to classify the DR grade presented in the images. Later, Section 5 shows first the metrics used to evaluate the

performance of the proposal and a sensitivity analysis of the scaling factor. Moreover, this section reports the watermarking performance over the whole fundus image dataset, the robustness against image processing and geometric attacks, a comparison of the proposed watermarking method versus other state-of-the-art works, and an evaluation of the DR diagnosis of both marked and non-marked images. Finally, Section 6 discusses the results of the present work, followed by the conclusions and future work given in Section 7.

### 2. Related Work

Storing patient information is becoming easier with digital medical records. These records can hold text data about the patient and their medical images, all in one secure digital file. One way to achieve this is by using digital watermarking, where patient information is hidden within the medical images themselves. However, this process must not alter the original medical images, as any changes could affect diagnosis. Therefore, a key measure of success for this technology is to guarantee the original diagnosis accuracy of the images after watermarking them.

In this section, first, we report the recent watermarking algorithms in the transform domain and those applied to fundus images. Secondly, we present the deep learning approaches used for DR classification in this kind of image.

## 2.1. Watermarking Algorithms

In [13], the authors describe a watermarking method to store and transmit digital fundus images with patient information. The authors used as a watermark the patient data. They calculated the histogram from the medical image, to determine a zero point and then a peak point because the number of pixels that are associated with the peak point is the number of bits that it is possible to insert. Klington et al. [14] developed a watermarking algorithm to authenticate the digital fundus images using SVD and DWT. As a watermark, the authors used textual information about the fundus images and the original image. According to the results, the maximum capacity to insert is 329,960 bits, and with the jittering attack, the algorithm modifies 43% of the total number of embedded pixels. The authors indicate that they used the green channel of the original image to generate the watermark, and the other channels (red and blue) were used to embed the watermark. In [12], Dey et al. presented a watermarking method applied to fundus images to insert an EPR into the blood vessels extracted using K-means segmentation. Additionally, the authors evaluated and found a percentage difference in accuracy (0.25%) in blood vessel extraction before and after watermarking. On the other hand, Singh et al. [15] reported a watermarking approach to insert a unique identification code into the blood vessels. The blood vessels are detected through the matched filter and derivative of Gaussian and the pattern found formed the personal identification code.

In other papers, the authors use all channels. For example, in the paper An Imperceptible Semi-blind Color Image Watermarking Using RDWT and SVD [16], the authors use the RGB color space and employ the three channels to insert the watermark. Also, they used Redundant Discrete Wavelet Transform (RDWT) and SVD. One of the advantages of this method is the amount of information to use as a watermark because the original image and the watermark are the same size. In [17], the authors used a hybrid method to design a robust and imperceptible watermark method for digital images. The authors combine Lifting Wavelet Transform (LWT), Schur Decomposition, and SVD. The watermark employed is a QR code of size  $256 \times 256$ . According to the results, the method has better robustness in comparison with other techniques. A novel hybrid watermarking algorithm is described in [10]. The authors present a brief review of the state-of-the-art and describe their proposal. It consists of a hybrid reversible-zero watermarking to verify the copyright and authenticity of medical images. This kind of scheme has the advantage of no distortion in the medical image because the copyright information is not directly embedded in it. The proposal is focused on a distortion-free method because in this way the accuracy of medical diagnosis is better. The results show that the watermark is recovered efficiently

and the authors compare their proposal with zero watermarking schemes and reversible watermarking algorithms.

In [18], the authors describe their method as the first algorithm in which ROI and RONI are divided only for watermark generation, not for watermark embedding. To embed the watermark they use Slantlet Transform (SLT)-SVD and Recursive Dither Modulation (RDM). Their results demonstrate a robust method for Average Filter, Gaussian Blurring, Gaussian Noise, JPEG Compression, Median Filter, Crop, Salt and Noise, Resizing, and Wiener Filter. Another hybrid method is [19]. In this paper, the authors present a method that uses Fast Discrete Curvelet Transform (FDCuT), DCT, SVD, and Arnold cat map. To increase security they employed the Arnold cat map because the original image is distorted to encrypt it and does not change its intensity. The hybrid method is focused on applications in medical images because of their imperceptibility and high resistance. They use different datasets to probe their algorithm, including [20,21]. The problem is the authors do not compare their method with other similar ones, even though their results are good in the insertion process and extraction process. The Peak Signal-to-Noise Ratio (PSNR) is close to 50 dB and the correlation is 1.00. A method to protect biometric images is presented in [22]. This method uses DWT, and SVD with Chaotic encryption. As a watermark, they use fingerprints and gait biometrics (20,000) and 1000 fundus images of different categories as original images, but also the authors probe their algorithm with common images. Their results are good in terms of imperceptibility because when using medical images PSNR = 50.43 dB, and when they used common images PSNR = 52.97 dB. The method was tested with different attacks: Gamma Correction, Resize, Rotate, Crop, Sharp, Speckle Noise, Adjust, and Salt and Pepper Noise, but they do not indicate the parameter values of each attack. Finally, they test their algorithm with different techniques to demonstrate their high imperceptibility.

#### 2.2. Deep Learning Classification Algorithms

On the other hand, there is a lot of research focused on DR detection using Artificial Neural Networks (ANN) or some variants, with some using different techniques. The state-of-the-art is very extensive because the authors try to develop the best technique for DR detection or some similar conditions, for example in [23], Kapoor and Arora presented a set of steps for DR detection and degree classification using deep CNNs. The authors applied the Enhance Local Contrast (CLAHE) method to obtain noise-free images from the Kaggle DR dataset [24], allowing the lesions to be visible, they classified the images using the Deep CNN. On the other hand, Radha et al. [25] presented a study of fundus eyes images, including the normalization of shape and size, segmentation, and automatic retinal lesion classification using the Atrous Convolutional Neural Network (ACNN) to extract the relevant features. Dutta et al. [26] proposed a feature extraction method from retinal images to perform binary and multi-class classification through various machine learning models. The authors used a variant of CNN using a transfer learning approach and hyper-parameter tuning of the VGG-19 model. Gayathri et al. [27] developed an automated DR grading method from the fundus images. A Multipath Convolutional Neural Network (M-CNN) was used for global and local feature extraction from images, and the machine learning classifiers Support Vector Machine (SVM), Random Forest, and J48 were used to categorize the input according to the severity. The model was evaluated across the publicly available databases IDRiD, Kaggle, and MESSIDOR. In [28], the authors Chetoui and Akhloufi presented a study to develop a deep learning algorithm capable of detecting DR on retinal fundus images. Also, the proposed deep learning algorithm fine-tunes a pre-trained deep CNN for DR detection. A new variant of fully convolutional networks, with its expansive path redesigned, is Lesion-Net [29]. A dual loss that leverages both semantic segmentation and image classification losses is introduced. Lastly, the authors have built a multi-task network that employs Lesion-Net as a side-attention branch for both DR grading and result interpretation. In [30], Ni et al. presented a deep convolutional neural network for DR stage classification, trained and evaluated. In addition, the model uses high-resolution retinal fundus images as inputs to take advantage of more detailed retinal lesion information in

images and a strong correlation between both eyes. Randive et al. [31], made a model that includes preprocessing, feature extraction using Spherical Directional Local Ternary Pattern (SDLTP), and classification using traditional distance measure and learning-based distance measure using ANN. Also, the SDLTP was used for extracting the directional feature in the 3D plane and to reduce the feature vector length. Loheswaran [32] developed a classification system using Fuzzy C-Means and a Recurrent Neural Network (RNN). Other proposals use different techniques or methods. In [33], Shorfuzzaman et al. presented an explainable deep learning ensemble model by fusing the weights from different models into a single model. It extracts salient features from various retinal lesions found on retinal fundus images. Then, the extracted features were fed to a custom classifier to obtain a DR severity level. The model was trained on the APTOS dataset and was tested using the APTOS, MESSIDOR, and IDRiD datasets. Suresh et al. [34] presented a screening technique that relies on the texture analysis of the retinal background using Local Ternary Patterns (LTP). Also, it compared the results obtained using the proposed approach with Local Binary Patterns (LBP) instead of LTP. They performed three experiments separating, DR from normal, Age-related Macular Degeneration (AMD) from normal, and DR from AMD. Sharif and Shah [35] presented an automatic design for the retinal lesions screening to grade the DR system. Also, the system is comprised of a preprocessing determination of biomarkers and formulation of a profile set for classification. In [36], Wang et al. presented a method requiring only a series of normal and abnormal retinal images without the need to specifically annotate their locations and types. Additionally, the proposed method encodes both the background knowledge of fundus images and the background noise into one unique model. On the other hand, Kaur and Mittal [37] presented a reliable segmentation of lesions that have been performed using iterative clustering irrespective of associated heterogeneity, and bright and faint edges. Afterwards, a computer-aided severity level detection method that they proposed to diagnose non-proliferative diabetic retinopathy. DelaPava et al. [38] designed a model for automatic DR classification on eye fundus images. The approach that has identifies the main ocular lesions related to DR and subsequently diagnoses the illness. Additionally, the Kaggle EyePACS subset is used as a training set, and the MESSIDOR-2 as a test set for lesions and DR classification models. A comprehensive machine learning computer-aided diagnosis (CAD) system based on deep learning techniques [39], eliminates noise, enhances quality, and standardizes the sizes of the retinal images. Also, it distinguishes between healthy and DR cases. Finally, the proposal automatically extracts the four changes: exudates, microaneurysms, hemorrhages, and blood vessels. Biswas et al. [40] developed a model called intelligent system for diabetic retinopathy for early detection of DR using a SVM.

## 3. Materials and Methods

In this section, the database of retinal fundus images used in the proposed watermarking method is introduced. Moreover, the fundamentals of the SHT, JT, and SVD are given. In addition, the theory of the CNN and the transfer learning strategy are explained in detail.

## 3.1. Fundus Image Dataset and Watermark Image

The proposed image watermarking and classification method of DR uses the image public dataset MESSIDOR-2 [20,21]. MESSIDOR stands for "Methods to Evaluate Segmentation and Indexing Techniques in the Field of Retinal Ophthalmology" (from french: Méthodes d'Évaluation de Systèmes de Segmentation et d'Indexation Dédiées à l'Ophtalmologie Rétinienne). The main objective of MESSIDOR project is to compare and evaluate segmentation algorithms for detecting lesions in retinal RGB images and to facilitate computer-assisted diagnoses of DR.

The MESSIDOR dataset is a collection of DR examinations, each one consisting of two macula-centered eye fundus RGB images. The MESSIDOR-original dataset was provided by the MESSIDOR program partners, containing 1058 images in PNG format (529 exami-

nations). The MESSIDOR-Extension dataset includes examinations from Brest University Hospital. It contains 690 images in JPEG format (345 examinations).

The MESSIDOR-2 dataset contains 1748 8-bit color images (874 examinations) of different sizes ( $1440 \times 960$  and  $2240 \times 1488$  pixels). It includes the MESSIDOR-original and the MESSIDOR-Extension. This dataset comes with a list containing image pairing and it does not contain annotations of DR. However, there are third parties that include these annotations. Currently, the MESSIDOR-2 dataset has errors from its site, and the dataset used in this work was downloaded from the Kaggle public dataset MESSIDOR-2: https://www.kaggle.com/datasets/geracollante/messidor2/, (accessed on 10 August 2023).

In addition, in [41], Krause et al. reported the DR grades (0, 1, 2, 3, 4) and Diabetic Macular Edema (DME) presence (0, 1) for the MESSIDOR-2 fundus image dataset, which is publicly available at: https://www.kaggle.com/datasets/google-brain/messidor2-drgrades, (accessed on 10 August 2023). The grades of the images were adjudicated by a panel of specialists. This is because the authors of the original MESSIDOR-2 dataset did not include any diabetic retinopathy ground truth. Figure 1 shows 25 examples of the MESSIDOR-2 fundus image dataset.



Figure 1. The 25 examples of the MESSIDOR-2 fundus image dataset.

As a watermark, we use the medicine symbol *Caduceus*, a Greek symbol representing the staff carried by the deity Hermes. The image of the symbol *Caduceus* used in this work is a gray-scale image of  $256 \times 256$  pixels as is shown in Figure 2:



Figure 2. Medicine symbol Caduceus used as watermark.

#### 3.2. Overview of the Proposed Bio-Inspired Watermarking Method

The proposed bio-inspired watermarking algorithm is shown in Figure 3. It is based on the Steered Hermite transform, Singular Value Decomposition, and Jigsaw transform. Our

proposal consists of an insertion process to embed a watermark image into the MESSIDOR-2 fundus image dataset and an extraction process to recover the watermark. In addition, a CNN model is used to estimate the DR grading. This parameter allows us to define if the original diagnosis is modified because of the watermark. We detail each method in the following sections.



Figure 3. Proposed hybrid watermarking algorithm.

The watermarking insertion process has as input the color MESSIDOR-2 fundus image dataset  $I_{RGB}(x, y, z)_i$  and the watermark W(x, y), where i = 1, ..., 1748 is the image index, x, y represent the spatial coordinates, and z = 1, 2, 3 the color band (Red, Green, Blue). Thus, the insertion process generates a set of watermarked fundus images  $I_{RGB}(x, y, z)_i$ . In addition, the watermarking process produces some elements, which are stored in the Key Area (yellow-shaded rectangle) and that will be used later in the extraction process.

On the other hand, the watermarking extraction process recovers the set of extracted watermarks  $\hat{I}_{RGB}(x, y, z)_i$ . Later, invisibility of the watermark, and robustness against attacks is performed over the watermarked dataset.

Finally, a CNN model is independently trained either with the original image dataset  $I_{RGB}(x, y, z)_i$  (dashed red line) or the watermarked image dataset  $\hat{I}_{RGB}(x, y, z)_i$  (dash-dotted blue line). Thus, both trained CNN models estimate the DR grading and are compared to evaluate whether the watermarking process modifies the DR diagnosis.

#### 3.3. Steered Hermite Transform

The Steered Hermite transform is obtained by rotating the image decomposition of the Hermite Transform (HT) [42], a bio-inspired image model used for a spatial-frequency decomposition of the digital image. Figure 4 shows the steps to calculate the SHT:



Figure 4. Steered Hermite transform schema.

First, the Cartesian Hermite coefficients  $(I_{m,n-m}(x, y))$  are obtained by convoluting a gray-scale image (2D image) (I(x, y)) with the Hermite analysis filters  $(D_{m,n-m})$ , as is shown in Equation (1), followed by sub-sampling with factor *T*.

$$I_{m,n-m}(x_0,y_0) = \sum_{(x_0,y_0)\in S} I(x,y) D_{m,n-m}(x_0-x,y_0-y),$$
(1)

where *m* and n - m represents the decomposition order in the spatial directions *x* and *y*, respectively, with  $m = 0 \dots n$  and  $n = 0 \dots D$ , *D* is the maximum order of the expansion, and  $(x_0, y_0)$  is the spatial position in the discrete sampling lattice *S*.

The Hermite analysis filters  $(D_{m,n-m})$  are obtained by the Equation (2):

$$D_{m,n-m}(x,y) = K_{m,n-m}(-x,-y)\omega^2(-x,-y),$$
(2)

where  $\omega^2(x, y)$  represents the 2D version of a binomial window function:

$$\omega(x) = \frac{1}{2^N} C_N^x, \quad x = 0, 1, \dots, N-1,$$

 $C_N^x$  is the binomial function:

$$C_N^x = \frac{N!}{(N-x)!} x!, \quad x = 0, 1, \dots, N-1,$$

 $K_{m,n-m}$  are the orthogonal polynomials associated with the binomial window  $\omega^2(x, y)$ :

$$K_n[x] = \frac{1}{\sqrt{C_N^n}} \sum_{k=0}^n (-1)^{n-k} C_{N-x}^{n-k} C_x^k,$$
(3)

and N + 1 is the size of the binomial window, and it is related to the size of the Gaussian window spread ( $\sigma$ ), so that defines the discrete sampling lattice *S*. Moreover, the maximum order of the expansion must fulfill the relationship  $D \le 2 * N$ .

Then, the steered Hermite coefficients  $(I_{n,\theta}(x, y))$  are calculated by rotating the Cartesian Hermite coefficients towards an angle  $\theta$  (Equation (4)):

$$I_{n,\theta}(x_0, y_0) = \sum_{k=0}^{n} \Big( I_{k,n-k}(x_0, y_0) \Big) \Big( \phi_{k,n-k}(\theta) \Big), \tag{4}$$

where  $\phi_{m,n-m}(\theta)$  are the angular functions, of order *n*, which are defined by Equation (5):

$$\phi_{m,n-m}(\theta) = \sqrt{\binom{n}{m}} \left( \cos^m(\theta) \right) \left( \sin^{n-m}(\theta) \right), \tag{5}$$

where  $\theta$  is an angle of maximum energy, e.g., the gradient angle.

The gradient angle could be approximated using the  $I_{0,1}(x, y)$  and  $I_{1,0}(x, y)$  Cartesian Hermite coefficients, which represent the edges in the horizontal and vertical directions of the image I(x, y), respectively. Equation (6) shows the approximation of the angle  $\theta$ :

$$\theta(x,y) = tg^{-1} \left[ \frac{I_{0,1}(x,y)}{I_{1,0}(x,y)} \right].$$
(6)

On the other hand, to reconstruct the original image the Inverse Steered Hermite Transform (ISHT) is applied as is shown in Figure 5:



Figure 5. Inverse Steered Hermite transform schema.

The Cartesian Hermite coefficients ( $I_{m,n-m}(x, y)$ ) are recovered by inversely rotating the steered Hermite coefficients ( $I_{n,\theta}(x, y)$ ), by applying Equations (4) and (6) with the stored angles ( $-\theta(x, y)$ ).

Next, the original image is reconstructed by calculating the Inverse Hermite Transform (IHT) through Equation (7). Before performing the reconstruction, an over-sampling with factor *T* is applied to the Cartesian Hermite coefficients.

$$I(x,y) = \sum_{n}^{N} \sum_{m=0}^{n} \sum_{(x_{0},y_{0})\in S} I_{m,n-m}(x_{0},y_{0}) P_{m,n-n}(x-x_{0},y-y_{0}),$$
(7)

where  $P_{m,n-n}$  are the Hermite synthesis filters given by Equation (8):

$$P_{m,n-m}(x,y) = \frac{D_{m,n-m}(x,y)}{W(x,y)},$$
(8)

and W(x, y) is a weight function:

$$V(x,y) = \sum_{(x_0,y_0) \in S} \omega^2(x - x_0, y - y_0) \neq 0$$

For multi-dimensional images (e.g., RGB images), the SHT is obtained by applying the HT to each band and then rotating the coefficients, generating a set of coefficients per band. Moreover, it is possible to decompose the whole color image using 3-D Hermite filters and then steering the Cartesian Hermite coefficients towards two local orientation angles ( $\theta$  and  $\phi$ ). See Mira et al. [43] for more details.

## 3.4. Jigsaw Transform

The Jigsaw transform is a popular scrambling technique used to hide visual information on digital images. It is calculated by relocating blocks of pixels of fixed size [44]. Thus, a digital image (I(x, y)) of  $X \times Y$  pixels could be divided into  $k \times l$  blocks of  $s_1 \times s_2$  pixels each one, so that  $k = X/s_1$  and  $l = Y/s_2$ . A random number generator defines the new location of the *j*-th block, with  $j = 1, ..., k \times l$ , and the original positions of each block are stored to recover the original image. Thus, the encryption/decryption process is symmetric and can be applied to both RGB and gray-scale images.

Figure 6 shows a fundus image of 960 × 11,140 pixels, and the JST results, applied to each color band of the RGB image, varying the number of blocks ( $k \times l$ ) and the size of each one ( $s_1 \times s_2$ ).



**Figure 6.** Examples of the JST a fundus image of 960 × 1140 pixels, the number of blocks (*N*) and the size of each one  $(s_1 \times s_2)$  were varied. (a) Fundus image. (b)  $k \times l = 96$  and  $s_1 \times s_2 = 120 \times 120$ . (c)  $k \times l = 384$  and  $s_1 \times s_2 = 60 \times 60$ . (d)  $k \times l = 1536$  and  $s_1 \times s_2 = 30 \times 30$ . (e)  $k \times l = 6144$  and  $s_1 \times s_2 = 15 \times 15$ . (f)  $k \times l = 1536$  and  $s_1 \times s_2 = 30 \times 30$ .  $k \times l = 1,382,400$  and  $s_1 \times s_2 = 1 \times 1$ .

#### 3.5. Singular Value Decomposition

Singular Value Decomposition is a mathematical tool used to determine the intrinsic algebra present in a matrix. For example, let *I* a gray-scale image of  $X \times Y$ , the SVD decomposed it into three matrices as is shown in Equation (9):

$$I = \sum_{i=1}^{r} \sigma_i u_i v_i^T = U S V^T, \tag{9}$$

where  $U \in \mathbb{R}^{Y \times Y}$ ,  $V \in \mathbb{R}^{X \times X}$  are orthogonal matrices and  $S \in \mathbb{R}^{X \times Y}$  is diagonal:

$$S = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & \sigma_r \end{bmatrix},$$
(10)

and the values  $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_r$  are unique, and are the singular values of *I*. The number  $r \le min(X, Y)$  is equal to the rank of *I*.

In image processing, SVD has been applied to several problems, such as compression, noise reduction, and watermarking [45]. One interesting property of SVD in images is related to the singular values, which do not change when the image is altered, for example, by image processing operations.

#### 3.6. Convolutional Neural Networks

Nowadays, the convolutional neural network is a novelty approach that has obtained efficient results in different fields of science and technology, such as computer vision, illness detection, image segmentation, self-driving, biometric authentication, and the entertainment industry.

The basis of CNNs is convolutional layers that extract local features (e.g., edges and textures) from a set of input images through several convolution kernels [46]. The 2-D kernels or filters are represented by matrices, and layers output is known as a feature map. Thus, each convolutional layer could be trained to detect determinate local features, generating a output feature map. In addition, pooling layers (or down-sampling layer) are used to reduce the dimension of the output of each convolutional layer, decreasing the

number of neurons and reducing in consequence overfitting and computational complexity. computational complexity. Max-pooling layer, min-pooling layer, and average-pooling layer are three classes of pooling layers, being the max-pooling layer the most commonly used. Figure 7 shows a classical architecture of a CNN, consisting of various convolutional-pooling layer pairs for feature extraction. Moreover, fully connected layers are used for classification tasks, and in the final process an activation function generated the output classification labels.



Figure 7. Classic architecture of a convolutional neural network.

The unknown parameters of a CNN model are the weights and biases of the connections. Thus, an iterative training process adjusts these weights for a later classification process [46]. In a typical training process, a high number of training samples are used. For example, ImageNet [47] is a dataset containing more than 14 million images in 1000 classes. However, in those cases where a large set of samples is not available, the transfer learning approach is an alternative. This method transfers the weights of a pre-trained network, with many samples, to the desired network to be trained by performing a fine-tuning [48]. Thus, the transfer learning replaces the last fully connected layer of the pre-trained CNN with a specific fully connected layer of the problem to be resolved [46].

For many classification problems, a fine-tuning of the last layers is performed, whereas the first layers remain without changes. However, if there is a significant difference between the data source of the pre-trained network and the data source of the current network, a fine-tuning of the first layers could contribute to extracting primitive features (gray levels, edges, colors, etc.) of the desired data samples [49].

Regarding pre-trained networks, VGG16, Inception-v3, and ResNet50 are three relevant networks used on medical image problems by applied transfer learning. These networks obtained the highest performing CNNs on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC2014) [50]. On the other hand, YOLO (You Only Look Once) is a real-time object detection algorithm that has gained popularity since 2016, when YOLOv2 was released [51].

On the one hand, VGG-16 is a network with 16 convolutional layers, two fullyconnected layers, a softmax classifier,  $3 \times 3$  convolutional filters, and  $2 \times 2$  max-pooling operations [52]. Inception-v3 CNN contains 48 convolutional layers and incorporates the so-called "Inception modules", which reduce the number of parameters while holding the network efficiency [53]. On the other hand, ResNet50 contains 50 convolutional layers into five blocks and it uses residual blocks by the so-called "identity shortcut connection" that skips one or more layers. The first block is composed of a convolutional and a pooling layer. The following four blocks use a layer stack of  $1 \times 1$ ,  $3 \times 3$ , and  $1 \times 1$  convolutional filter size, respectively, with three, four, six, and three layers for each block.

In addition, YOLOv8 from Ultralytics [54] incorporates anchor boxes from Faster R-CNN, which are used to size and shape real-time prediction of the objects [51]. It has two main parts: the backbone and the head. The backbone of YOLOv8 is a modified version of the CSPDarknet53 architecture, consisting of 53 convolutional layers that use cross-stage partial connections to improve information flow between the layers. The head of YOLOv8 consists of multiple convolutional layers followed by a series of fully connected layers.

These layers are responsible for predicting bounding boxes, objectness scores, and class probabilities for the objects detected in an image. Finally, YOLOv8 can perform multi-scaled object detection using a feature pyramid network to detect objects of different sizes and scales within an image.

#### 4. Proposed Method for Image Watermarking

The present article is an extension of a previous work published in [55]. In that work, we proposed a hybrid watermark method based on the SHT, SVD, and JST. The original algorithm uses the SHT because it is possible to insert the watermark in oriented structures. This has the advantage that the watermark is more imperceptible to the Human Vision System (HVS). On the other hand, we employ SVD because we ensure more robustness to different attacks using the second level of SVD. In addition, the algorithm was tested with chest X-ray gray-scale images extracted from the Kaggle public dataset COVID-19 Radiography Dataset (https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database/, (accessed on 14 February 2022)). The results showed that medical images do not suffer visual alterations reporting PSNR values up to 30 dB, Mean Structure Similarity Index (MSIM) values are within 0.9800, and Normalized Crossed-Correlation (NCC) values equal to 0.9900.

In this paper, we propose to use a variation of the algorithm presented in [55], applied to RGB fundus images, and evaluate whether the watermark insertion process modifies the DR diagnosis.

Thus, the main differences with the previous work are listed as follows: (*i*) The proposed watermarking method was adapted and tested on RGB fundus images instead of gray-scale images, as was reported in [55]. (*ii*) The fundus images were first transformed into the Y'CbCr color model, then the luma channel was selected to insert the watermark, and the SHT for 2D images (Equations (1) and (4)) was applied. (*iii*) The watermark was embedded into the blood vessel patterns extracted by the SHT in the luma channel, obtaining in consequence a natural and automatic ROI-based watermarking method, and differing from [55] where the watermark is embedded in the whole image. (*iv*) The encryption security of the watermark was increased by using blocks of  $1 \times 1$  of the JST. (*v*) The watermarking method was tested over 1748 8-bit color images of different sizes (1440 × 960 and 2240 × 1488 pixels). (*vi*) A complete robust analysis was performed by applying 11 different attacks instead only of six attacks from the previous work. (*vii*) A classification task using different deep learning architectures was carried out to identify diagnosis changes in the watermarked images. Finally, (*viii*), with the proposed modifications we overcome the evaluation metrics of the previous work.

# 4.1. Watermarking Insertion Process

Figure 8 shows the details of the proposed watermarking insertion schema using the steered Hermite transform and the SVD technique. The solid blue line corresponds to steps for the color model conversion of the cover fundus image, its spatial-frequency decomposition using the SHT and the SVD, and the watermarking embedding step. On the other hand, the dashed red line represents the following steps for pre-processing and encrypting the watermark using the JST. In addition, the green dotted line shows steps to reconstruct the watermarked fundus image using the inverse SVD, the ISHT, and the color model conversion. Finally, the yellow-shaded ellipses correspond to those elements stored in the Key Area used in the extraction process, and the symbols \* and + represent the multiplication and addition operations, respectively.



Figure 8. Watermarking insertion schema.

Thus, the steps of embedding the watermark (W(x, y)) are described below:

- 1. Input the RGB fundus image  $I_{RGB}(x, y, z)$ .
- 2. Convert the fundus image  $I_{RGB}(x, y, z)$  from the RGB model to the Y'CbCr (Y represents the luma component or the brightness in an image, Cb and Cr are the bluedifference and red-difference chrome components, respectively). In the Equation (11) we present this conversion as is described in [56]:

$$\begin{bmatrix} I_{Y'}(x,y)\\ I_{Cb}(x,y)\\ I_{Cr}(x,y) \end{bmatrix} = \begin{bmatrix} 16\\ 128\\ 128 \end{bmatrix} + \begin{bmatrix} 65.481 & 128.553 & 24.966\\ -37.797 & -74.203 & 112\\ 112 & -93.786 & -18.214 \end{bmatrix} \cdot \begin{bmatrix} I_R(x,y)\\ I_G(x,y)\\ I_B(x,y) \end{bmatrix}, \quad (11)$$

where  $I_R(x, y)$ ,  $I_G(x, y)$  and  $I_B(x, y)$  correspond to the color bands of the image  $I_{RGB}(x, y)$ , and  $I_{Y'}(x, y)$ ,  $I_{Cb}(x, y)$  and  $I_{Cr}(x, y)$  correspond to the components of the image  $I_{Y'CbCr}(x, y)$ .

- 3. Select the luma band  $I_{Y'}(x, y)$  of  $I_{Y'CbCr}(x, y)$ . We found that the Y' component is the band most suitable to embed the watermark.
- 4. Calculate the SHT for 2D images (Equations (1) and (4)) to the luma band  $I_{Y'}(x, y)$ , obtaining the steered Hermite coefficients  $I_{n,\theta}(x, y)$ . The parameters used are N = 2, D = 2 \* N, and T = 2.
- 5. Select the steered Hermite coefficient  $I_{i,\theta}(x, y)$  for embedding the watermark.
- 6. Apply the SVD to  $I_{j,\theta}(x, y)$  obtaining the matrices  $U_1(x, y)$ ,  $S_1(x, y)$  and  $V_1(x, y)$ . Store the diagonal matrix of singular values  $S_1(x, y)$  into the Key Area.
- 7. Input the watermark image W(x, y).
- 8. Resize W(x, y) to have the same size of  $I_{i,\theta}(x, y)$ .
- 9. Encrypt the resized matrix W(x, y) by applying the JST, by dividing the resized watermark into  $k \times l = X \times Y$  blocks of  $s_1 \times s_2 = 1 \times 1$  pixels each one, and obtaining the matrix  $W_{IST}(x, y)$ . The original indexes Idx of the blocks are stored in the Key Area.
- 10. Embed the encrypted watermark  $W_{IST}(x, y)$  in the matrix  $S_1(x, y)$  applying the Equation (12):

$$S_W(x,y) = S_1(x,y) + \alpha * W_{JST}(x,y),$$
 (12)

where  $\alpha$  is a scaling factor, which defines the imperceptibility of the watermark image.

- 11. Apply a second SVD to  $S_W(x, y)$  abstaining the matrices  $U_2(x, y)$ ,  $S_2(x, y)$  and  $V_2(x, y)$ . Store the left and right singular matrices  $U_2(x, y)$  and  $V_2(x, y)$  into the Key Area.
- 12. Apply the inverse SVD using the matrices  $S_2(x, y)$ ,  $U_1(x, y)$  and  $V_1(x, y)$  obtaining the marked steered Hermite coefficient  $\hat{I}_{i,\theta}(x, y)$ .
- 13. Perform the assembly of the steered Hermite coefficients using the coefficients  $I_{n,\theta}(x, y)$  (step 4) and including the marked coefficient  $\hat{I}_{i,\theta}(x, y)$  (step 12).
- 14. Calculate the ISHT (Equation (4) with  $\theta(x, y)' = -\theta(x, y)$  and Equation (7)), obtaining the marked luma band  $\hat{I}_{Y'}(x, y)$ .
- 15. Finally, using the bands  $I_{Cb}(x, y)$  and  $I_{Cr}(x, y)$  of  $I_{Y'CbCr}(x, y, z)$  (step 2) and the marked luma band  $\hat{I}_{Y'}(x, y)$ , convert from the Y'CbCr color model to the RGB model obtaining the marked fundus image  $\hat{I}_{RGB}(x, y, z)$ . Equation (13) show this conversion [56]:

$$\begin{bmatrix} I_R(x,y)\\ I_G(x,y)\\ I_B(x,y) \end{bmatrix} = \begin{bmatrix} 0.00456621 & 0 & 0.00625893\\ 0.00456621 & -0.00153632 & -0.00318811\\ 0.00456621 & 0.00791071 & 0 \end{bmatrix} \cdot \left( \begin{bmatrix} I_{Y'}(x,y)\\ I_{Cb}(x,y)\\ I_{Cr}(x,y) \end{bmatrix} - \begin{bmatrix} 16\\ 128\\ 128 \end{bmatrix} \right).$$
(13)

An important factor in the proposed watermarking method corresponds to the selection of the steered Hermite coefficient  $I_{j,\theta}(x, y)$  for embedding the watermark because is important to have robustness and imperceptibility. So, we performed several tests and we selected the coefficient  $I_{2,\theta}(x, y)$ . Figure 9 shows the luma component  $I_{Y'}(x, y)$  (Figure 9a) of a fundus image and its corresponding coefficient  $I_{2,\theta}(x, y)$  (Figure 9b).



(a)



(b)

**Figure 9.** (a) Luma component  $(I_{Y'}(x, y))$ . (b) Selected Steered Hermite coefficient  $(I_{2,\theta}(x, y))$ .

The selected coefficient allows for inserting the watermark into the blood vessel patterns, giving a natural and automatic ROI-based watermarking method. Thus, the proposed method is an alternative to Dey et al. [12], where the blood vessels are first extracted using K-means segmentation and then the EPR is hidden into them using interpolation and trigonometric functions. Also, our approach differs from the work of Singh et al. [15], where a unique identification code is inserted into the blood vessels detected in fundus images. In that work, the blood vessels are detected through the matched filter and derivative of Gaussian, and these patterns, unique per patient, represent the identification code.

#### 4.2. Watermarking Extraction Process

In Figure 10, we can see the proposed watermarking extraction schema. It is composed of two stages. First, the solid blue line corresponds to the decomposition of the marked fundus image using the SHT and SVD, followed by the operations to extract the watermark. Second, the dashed red line represents the decryption steps, using the inverse JST and the post-processing operations to recover the watermark. On the other hand, the yellow-shaded



ellipses are those elements stored in the Key Area in the insertion phase and used to recover the watermark.

Figure 10. Watermarking extraction schema.

The extraction process is described below:

- 1. Input the marked fundus image  $\hat{I}_{RGB}(x, y)$ .
- 2. Convert the marked fundus image  $\hat{I}_{RGB}(x, y)$  to the Y'CbCr color model (Equation (11)), obtaining the image  $\hat{I}_{Y'CbCr}(x, y)$ .
- 3. Select the luma band of  $\hat{I}_{Y'CbCr}(x, y)$ .
- 4. Calculate the SHT for 2D images (Equations (1) and (4)) to the luma band  $\hat{l}_{Y'}(x, y)$ , with the same parameters employed in the insertion process, obtaining the steered Hermite coefficients  $\hat{l}_{n,\theta}(x, y)$ .
- 5. Select the steered Hermite coefficient  $\hat{I}_{i,\theta}(x, y)$  to extract the watermark.
- 6. Calculate the SVD to  $\hat{I}_{j,\theta}(x,y)$  obtaining the matrices  $\hat{U}_1(x,y)$ ,  $\hat{S}_1(x,y)$  and  $\hat{V}_1(x,y)$ .
- 7. Calculate a second SVD to the matrix  $\hat{S}_1(x, y)$  obtaining the matrices  $\hat{U}_2(x, y)$ ,  $\hat{S}_2(x, y)$  and  $\hat{V}_2(x, y)$ .
- 8. Read the left and right singular matrices  $U_2(x, y)$  and  $V_2(x, y)$  from the Key Area (yellow-shaded rectangle).
- 9. Apply an inverse SVD decomposition using matrices  $\hat{S}_2(x, y)$ ,  $U_2(x, y)$ , and  $U_2(x, y)$  to obtain the diagonal matrix of singular values  $\hat{S}_W(x, y)$ .
- 10. Read the diagonal matrix of singular values  $S_1(x, y)$  from the Key Area (yellow-shaded rectangle).
- 11. To extract the first version of the watermark we apply the Equation (14):

$$\hat{W}_{JST}(x,y) = \frac{\hat{S}_W(x,y) - S_1(x,y)}{\alpha}.$$
(14)

12. Read the indexes *Idx* from the Key Area (yellow-shaded rectangle) used in the JST of the insertion process.

- 13. Apply the inverse Jigsaw transform to  $\hat{W}_{JST}(x, y)$ , using the indexes Idx, to obtain the original position of the pixels, and recover the watermark image.
- 14. Finally, the extracted watermark is resized to its original dimensions to obtain the final watermark image  $\hat{W}(x, y)$ .

#### 4.3. Network Architectures for DR Classification

For the DR classification of the images present in MESSIDOR-2 dataset we used the CNN architectures VGG16, ResNet50, and InceptionV3 through Tesorflow-Keras [57,58]. In addition, we used the YOLOv8 network due to its simplicity, speed, and accuracy through the implementation of Ultralytics [54].

As to VGG16, ResNet50, and InceptionV3, we used their standard Keras Implementation as defined in the "tf.keras.applications" module [58] and as it is shown in Table 1:

Network	Number of Layers	Type of Filters	<b>Residual Connections</b>
VGG16	16	$3 \times 3$	No
InceptionV3	48	$1 \times 1$ , $3 \times 3$ , $5 \times 5$	No
ResNet50	50	$3 \times 3$	Yes

Table 1. VGG16, ResNet50, and InceptionV3 CNN implementation.

For the YOLOv8 architecture, we selected the YOLOv8n variant, which is the smallest and simplest one.

The MESSIDOR-2 dataset was divided into two parts: 80% for training and 20% for testing. Tools like Keras and YOLO offer straightforward methods to create classes without manual annotation. However, data should be organized in a specific file structure.

The root directory contains sub-folders named with class identifiers (0, 1, 2, ...). Each sub-folder corresponds to a class and contains images (img1.jpg, img2.jpg, ...) belonging to that class. This structure is effective as each image features a retinal image and is associated with only one class:

```
root/
|-0/
| | – img1.jpg
| | – img2.jpg
| |-...
|-1/
| |- img3.jpg
| |- img4.jpg
| |-...
|-2/
| | – img5.jpg
| | – img6.jpg
| |-...
Τ
I-...
```

Regarding the classification categories, we utilized the "adjudicated DR Grade" from the MESSIDOR-2 DR Grades dataset [41] as the basis for our classes. This classification is crucial for evaluating how watermarking affects the neural network's ability to correctly identify classes.

As per described previously, the categories used for classification (adjudicated\_dr\_grade) are as follows, according to the five-point International Clinical Diabetic Retinopathy

(ICDR) grade:

- 0 = None
- 1 = Mild DR
- 2 = Moderate DR
- 3 =Severe DR
- 4 = PDR

# 5. Experiments and Results

# 5.1. Experimental Setup

We implemented the proposed watermarking algorithm in a Dell laptop Inspiron 7380 (Dell, Inc., Round Rock, TX, USA) on an Intel Core i7 @ 1.6 GHz (Intel Corporation, Santa Clara, CA, USA) with 16 GB RAM. The method was programmed in a non-optimized script in MATLAB R2018b (MathWorks, Natick, MA, USA) without a parallel configuration. The parameters used were: (*i*) For the SHT an analysis window size N = 2, a maximum expansion order D = 2 \* N = 4, and a sub-sampling factor T = 2. (*ii*) Blocks of size  $s_1 \times s_2 = 1 \times 1$  pixels for the JST. (*iii*) A scaling factor  $\alpha = 1 \times 10^{-5}$  (see Section 5.3).

Alternatively, the CNN architectures were developed on a personal custom built computer operating with Manjaro Linux KDE Edition (Manjaro GmbH & Co. KG, Grafing b. München, Bavaria, Germany). This system was equipped with a Ryzen 5 3600 processor (Advanced Micro Devices, Inc., Santa Clara, CA, USA), an NVIDIA GeForce RTX 3070 8 GB graphics card (Nvidia Corporation, Santa Clara, CA, USA), and 32 GB of RAM.

The watermarking algorithm has a time-consumption of 0.67061 s for the insertion stage and 0.78226 s for the extraction using images of  $2240 \times 1488$  pixels. These times were obtained without the JST step, which is the stage that consumes more time when small size blocks are used ( $1 \times 1$ ,  $2 \times 2$ ,  $3 \times 3$ , etc.).

# 5.2. Evaluation Metrics

In this section, we describe two sets of evaluation metrics. First, we reported the watermarking metrics to measure the quality of the watermarked images and the extracted watermark. Secondly, we presented the metric used to test the influence of the watermark on the diagnosis of DR.

#### 5.2.1. Watermarking Metrics

To evaluate the algorithm we use different metrics that are commonly used in watermarking. Some of them are focused on evaluating image quality and others on determining if the image has suffered modifications compared with the original. Considering the images  $I_1$  (original image),  $I_2$  (watermarked image), with  $X \times Y$  size, and x and y represent the spatial coordinates we define the metrics as follow:

• Mean Square Error (MSE) is a quality criterion in image processing. It evaluates the similarity between two images (*I*<sub>1</sub>, *I*<sub>2</sub>). See Equation (15):

$$MSE = \frac{1}{XY} \sum_{x=1}^{X} \sum_{y=1}^{Y} \left[ I_1(x, y) - I_2(x, y) \right]^2.$$
(15)

• PSNR is an objective metric to compare two images, using numerical criteria [59]. In Equation (16) it is defined:

$$PSNR = 10\log_{10}\left(\frac{255^2}{MSE}\right).$$
(16)

Close values to zero of MSE and higher values of PSNR correspond to a better quality of the watermarked image.

• NCC has been used as a metric to evaluate the degree of similarity between two images. Compared with ordinary cross-correlation, it is less sensitive to linear changes in the amplitude of illumination in two images [60]. We calculated it using Equation (17).

$$NCC = \frac{\sum_{x}^{X} \sum_{y}^{Y} (I_{1}(x,y) - \bar{I}_{1}) (I_{2}(x,y) - \bar{I}_{2})}{\left( \left[ \sum_{x}^{X} \sum_{y}^{Y} (I_{1}(x,y) - \bar{I}_{1})^{2} \right] \left[ \sum_{x}^{X} \sum_{y}^{Y} (I_{2}(x,y) - \bar{I}_{2})^{2} \right] \right)^{1/2}},$$
(17)

where  $\bar{I}_*$  represents the average value of  $I_*$ .

Close to one value of NCC corresponds to a watermarked image of better quality.

 Structural Similarity Index (SSIM) is a quality metric to measure the similarity between two images. It is considered to be correlated with the quality perception of the HVS [59]. It takes into account three factors: loss of correlation, luminance distortion, and contrast distortion. It is defined by Equation (18).

$$SSIM(I,\hat{I}) = \frac{(2\mu_I\mu_{\hat{I}} + C_1)(2\sigma_{I\hat{I}} + C_2)}{(\mu_I^2 + \mu_{\hat{I}}^2 + C_1)(\sigma_I^2 + \sigma_{\hat{I}}^2 + C_2)},$$
(18)

where *I* and  $\hat{I}$  correspond to the original and the distorted image, respectively. The averages of *I* and  $\hat{I}$  are given by  $\mu_I$  and  $\mu_{\hat{I}}$ , respectively, and their standard deviations are given by  $\sigma_I$  and  $\sigma_{\hat{I}}$ , the covariance between both images is represented by  $\sigma_{I\hat{I}}$ , and the constants  $C_1$  and  $C_2$  are used to prevent instability if the denominator happens to have a value close to zero.

MSSIM represents the SSIM mean. Is calculated as we show in Equation (19).

$$MSSIM(I, \hat{I}) = \frac{1}{M} \sum_{j=1}^{M} SSIM(I_j, \hat{I}_j),$$
(19)

where  $I_j$  and  $\hat{I}_j$  represent their *j*-th local window, and *M* stands for the number of local windows of the image. In the case that  $I_j$  and  $\hat{I}_j$  have no negative values. A better quality of the watermarked image is achieved for close to one value of SSIM

and MSSIM metrics.

# 5.2.2. Classification Metrics

To evaluate how the watermark embedding process affects the DR diagnosis in the fundus images, the accuracy classification metric was used as a support metric for the tests in all architectures (YOLOv8, VGG16, InceptionV3, and ResNet50), as it is shown in Equation (20):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN'}$$
(20)

where *TP*, *TN*, *FP*, and *FN* correspond to the True Positive, True Negative, False Positive, and False Negative values, respectively.

Accuracy is measured within the range [0, 1], with 0 as the worst and 1 as the best possible score.

### 5.3. Sensitivity Analysis of the Scaling Factor

The scaling factor  $\alpha$ , used in the insertion and extraction process, defines the imperceptibility of the watermark and its robustness against attacks. In this sense, low values give imperceptibility but decrease robustness. On the other hand, high values of the scaling factor provide robustness but affect the quality of the watermarked image. Therefore, a sensitivity analysis of the scaling factor was performed by marking with different values of  $\alpha$  a set of ten images from the image dataset and randomly selected. The values of the scaling factor varied from  $1 \times 10^{-6}$  to  $1 \times 10^{-4}$  with steps of  $1 \times 10^{-6}$ , generating 100 different results per image. Thus, we marked a total of 1000 images and we calculated the watermarking metrics reported in Section 5.2. Table 2 shows the average metrics obtained over the selected image set by varying the scale factor.

Table 2. Average performance metrics for the extracted watermark by varying the scale factor.

MSE	PSNR (dB)	NCC	SSIM	MSSIM
$4.87209  imes 10^{-6}$	53.50373	0.99936	0.99330	0.99354

The metrics shown in Table 2 mean that the scale factor  $\alpha$  does not affect the quality of the watermarked image in the insertion process.

On the other hand, Figure 11 shows the PSNR and the SSIM metrics of each selected image (A–J images) by varying the scale factor  $\alpha$ . In addition, we graph the threshold scale factor  $\alpha_{th} = 1 \times 10^{-5}$  (vertical black dotted line) which shows that a scale factor below  $\alpha_{th}$ , the PSNR value (Figure 11a) of some images decrease. In the case of the SSIM (Figure 11b), the same behavior occurs for values below  $\alpha_{th}$ .



**Figure 11.** Sensitivity analysis of the scaling factor  $\alpha$ , for the ten selected images (A–J images), by varying the scale factor. (a) PSNR values. (b) SSIM values.

In Figure 12, we show the images used for sensitivity analysis of Figure 11 (A–J images). We can see that for images B, C, D, G, I, and J, the PSNR and SSIM values remain constants for the values range of  $\alpha$ . In addition, for images A and E metrics increase and then hold constant for values of  $\alpha \ge \alpha_{th}$ . On the other hand, for images F and H, The metrics show higher values for  $\alpha \le \alpha_{th}$  and lower performance for values  $\alpha_{th} \le \alpha \le 5 \times 10^{-5}$ , which corresponds to those images where blood vessel are not visible or are very subtle. Thus, we fixed the scaling factor to  $\alpha = 1 \times 10^{-5}$  for all images of the dataset, which ensures both a high quality of the marked images and a successful extraction of the watermark.



Figure 12. Images used for sensitivity analysis of Figure 11: images (A–J).

## 5.4. Watermarking Performance Analysis

The proposed watermarking method was tested using the complete image public dataset MESSIDOR-2 [20,21]. The *Caduceus* symbol (Figure 2) was embedded following the steps presented in Section 4.1. Below we present the performance measures of the proposed method.

Table 3 presents the average performance metrics for the watermarked images (insertion stage) and the recovered watermarks (extraction stage).

**Table 3.** Average performance metrics using the 1748 fundus images of the MESSIDOR-2 dataset. The insertion section corresponds to the watermarked images and the extraction section to the recovered watermarks.

Insertion				Extraction					
MSE	PSNR (dB)	NCC	SSIM	MSSIM	MSE	PSNR (dB)	NCC	SSIM	MSSIM
$4.6976  imes 10^{-6}$	53.8638	0.9993	0.9937	0.9938	0.0007	32.0690	0.9975	0.9937	0.9943

On the other hand, Figure 13 shows some of the best examples of the insertion and extraction process. The left column corresponds to the original fundus images. The middle columns show the watermarked images and the right column shows the extracted watermarks, with their PSNR value, respectively. As we can see, these results show a good balance between the invisibility of the watermarks and the quality of the watermarked images, with high PSNR values of the watermarked images, and the quality of the extracted watermarks with high PSNR values.





In addition, Figure 14 shows some results with worse performance in the extracted watermark process, but with high metrics in the watermarked image. Thus, the original

fundus images are shown in the left column. The middle column shows the watermark images and the right column shows the extracted watermarks with their PSNR values, respectively. In these results, the PSNR values of the extracted watermarks decrease a bit. However, the high PSNR values of the watermarked images demonstrate that the invisibility of the watermarks and the quality of the watermarked images remain unchanging compared with the best results.



**Figure 14.** Worst results examples. Original images: (**a**,**d**,**g**). Watermarked images: (**b**,**e**,**h**). Recovery watermark: (**c**,**f**,**i**).

Comparing Figures 13 and 14, we can see that visually original and watermark images do not suffer alterations, and even though recovery watermarks in Figure 14 have PSNR values from 20 to 30 dB the images are clear. Therefore, a critical point studied and reported later is to determine if the DR grade diagnosis is modified in the watermarked images respecting the originals.

#### 5.5. Watermarking Robustness against Attacks

To measure the robustness of the present watermarking approach, we applied different attacks to the watermarked images of the complete dataset. To this aim, we vary the parameter that defines the operation in each attack. Following, in Table 4 we list the type of attack, the operation name, and the corresponding parameter that was applied:

Attack Type	Operation Name	Parameter Name	Parameter Value
	Gaussian Filter (GF)	Filter size	N  imes N
	Median Filter (MF)	Window size	$N \times N$
	Gaussian Noise (GN)	Variance	σ
Image Processing	Salt and Pepper Noise (SP)	Noise density	d
image r locessing	Contrast Enhancement (CE)	Percent saturation	P <sub>CE</sub> (%)
	Histogram Equalization (EQ)	Equalization Levels number	LV
	JPEG Compression (JPEGC)	Quality Percentage	QP (%)
	Image Scaling (SC)	Scaling factor	SF
Geometric	Rotation (ROT)	Rotation angle	$\phi$ (°)
	Cropping (CROP)	Cropping percentage	$P_{CR}$ (%)
	Translation (TRAN)	Displaced pixels number	$\Delta x, \Delta y$

Table 4. Definition of the attacks applied and their corresponding parameters.

In Table 5, we show the average results of the completed dataset by applying the following image processing operations to the watermarked image: Gaussian filter, Median Filter, Gaussian Noise, and Salt and Pepper Noise.

**Table 5.** Average performance metrics for the extracted watermark using the 1748 fundus images of the MESSIDOR-2 dataset and applying the attacks: GF, MF, GN, and SP.

Attack/Parameter	Parameter Value	MSE	PSNR (dB)	NCC	SSIM	MSSIM
	$3 \times 3$	$6.9476\times10^{-4}$	32.14117	0.99766	0.99409	0.99457
	$5 \times 5$	$7.2134 imes10^{-4}$	32.06983	0.99757	0.99383	0.99442
$GF/(N \times N)$	$7 \times 7$	$3.4721 imes10^{-3}$	30.06335	0.98959	0.96675	0.97031
	$9 \times 9$	$2.2502 imes10^{-1}$	18.69012	0.71425	0.66672	0.67248
	Average	$5.7477  imes 10^{-2}$	28.24112	0.92477	0.90535	0.90795
	$2 \times 2$	$5.7127\times10^{-4}$	32.69945	0.99808	0.99533	0.99534
	$3 \times 3$	$5.5048 imes10^{-4}$	32.83169	0.99815	0.99551	0.99550
$MF/(N \times N)$	4 imes 4	$5.5852 imes10^{-4}$	32.78804	0.99812	0.99545	0.99544
	$5 \times 5$	$5.9831 imes10^{-4}$	32.77550	0.99800	0.99501	0.99502
	Average	$5.6965 imes10^{-4}$	32.77367	0.99809	0.99532	0.99532
	0.10	$2.3761\times10^{-3}$	31.78001	0.99546	0.99108	0.99191
	0.20	$3.1808 imes10^{-3}$	31.59209	0.99437	0.98949	0.99052
$GN/(\sigma)$	0.40	$3.2338 imes10^{-3}$	31.44154	0.99417	0.98885	0.99010
	0.50	$3.2449\times10^{-3}$	31.40980	0.99413	0.98869	0.98998
	Average	$3.0089  imes 10^{-3}$	31.55586	0.99453	0.98952	0.99063
SP/( <i>d</i> )	0.20	$3.2188\times10^{-3}$	31.49094	0.99422	0.98898	0.99016
	0.40	$3.2418 imes10^{-3}$	31.43030	0.99416	0.98877	0.99002
	0.60	$3.2713\times10^{-3}$	31.37093	0.99407	0.98852	0.98987
	0.75	$3.2793  imes 10^{-3}$	31.33329	0.99404	0.98840	0.98979
	Average	$3.2528\times10^{-3}$	31.40636	0.99412	0.98867	0.98996

In Table 6, we show the average results of the completed dataset by applying the following image processing operations to the watermarked image: Contrast Enhancement, Histogram Equalization, JPEG Compression, and Image Scaling.

Attack/Parameter	Parameter Value	MSE	PSNR (dB)	NCC	SSIM	MSSIM
	0.5%	$3.5275\times10^{-2}$	28.50771	0.94652	0.91186	0.91404
	1.0%	$2.9811\times10^{-2}$	29.82147	0.95742	0.93547	0.93700
$CE/(P_{CE})$	2.0%	$4.4006 imes10^{-2}$	29.35476	0.93853	0.91139	0.91290
	3.0%	$4.0111 imes10^{-2}$	29.32348	0.94157	0.90969	0.91140
	4.0%	$3.6314\times10^{-2}$	29.48612	0.94682	0.91629	0.91815
	Average	$3.7104  imes 10^{-2}$	29.29871	0.94617	0.91694	0.91870
	8	$6.6205\times10^{-4}$	32.06528	0.99777	0.99460	0.99466
	32	$7.5019 imes10^{-4}$	31.51311	0.99747	0.99392	0.99397
EQ/(LV)	64	$7.6718 imes10^{-4}$	31.41761	0.99742	0.99378	0.99384
	128	$7.7192 imes10^{-4}$	31.39045	0.99740	0.99374	0.99380
	256	$7.7103 imes10^{-4}$	31.39286	0.99740	0.99375	0.99381
	Average	$7.4448 imes10^{-4}$	31.55586	0.99749	0.99396	0.99402
	90%	$8.5412\times 10^{-4}$	31.39271	0.99713	0.99252	0.99359
	80%	$2.4153\times10^{-2}$	24.04904	0.94505	0.88227	0.89217
IPECC/(OP)	70%	$9.7825  imes 10^{-2}$	19.76781	0.83312	0.71775	0.72688
JI LOC/ (QI )	60%	$1.5889 imes10^{-1}$	18.57334	0.76277	0.65339	0.66094
	50%	$1.5515 imes10^{-1}$	17.74861	0.76366	0.64656	0.65428
	Average	$8.7375  imes 10^{-2}$	22.30630	0.86034	0.77850	0.78557
	0.25  imes	$8.1853\times 10^{-1}$	0.86967	0.00346	0.13062	0.13400
SC/(SF)	0.50  imes	$5.9573  imes 10^{-1}$	5.24181	0.39417	0.26875	0.27398
	$1.50 \times$	$1.8756  imes 10^{-3}$	30.87795	0.99413	0.98253	0.98540
	$1.75 \times$	$2.1085\times10^{-3}$	30.73933	0.99347	0.98017	0.98318
	2.00  imes	$3.9879\times10^{-3}$	30.09281	0.98853	0.96374	0.96752
	Average	$2.8445  imes 10^{-1}$	19.56431	0.67475	0.66516	0.66882

**Table 6.** Average performance metrics for the extracted watermark using the 1748 fundus images of the MESSIDOR-2 dataset and applying the attacks: CE, EQ, JPEG, and SC.

In Table 7, we present the average results of the completed dataset by applying the following geometric attacks: Rotation, Cropping, and Translation.

Table 7. Av	verage performance	metrics for the extract	ed watermark ı	using the 1748	fundus in	mages of
the MESSIE	OOR-2 dataset and a	applying the attacks: R	OT, CROP and	l TRAN.		

Attack/Parameter	Parameter Value	MSE	PSNR (dB)	NCC	SSIM	MSSIM
	5°	$9.9763\times10^{-4}$	31.91858	0.99669	0.99097	0.99239
	$15^{\circ}$	$1.0727 imes10^{-3}$	31.87668	0.99649	0.99033	0.99181
	$45^{\circ}$	$1.1402 imes10^{-3}$	31.52496	0.99628	0.98954	0.99103
$ROT/(\phi)$	$65^{\circ}$	$7.9339 imes10^{-4}$	32.11663	0.99734	0.99301	0.99383
	$90^{\circ}$	$1.4098 imes10^{-3}$	28.80135	0.99527	0.98668	0.99021
	$190^{\circ}$	$7.0773 imes10^{-4}$	32.11481	0.99763	0.99391	0.99438
	Average	$\textbf{1.0202}\times \textbf{10}^{-3}$	31.39217	0.99662	0.99074	0.99227
	10%	$5.3050\times10^{-4}$	33.01504	0.99821	0.99569	0.99565
	20%	$5.3289 imes10^{-4}$	33.00432	0.99821	0.99567	0.99563
	30%	$5.5777 imes10^{-4}$	32.78566	0.99812	0.99545	0.99545
$CROP/(P_{CR})$	40%	$1.4770  imes 10^{-3}$	28.49738	0.99505	0.98599	0.98966
	50%	$9.0257 imes10^{-3}$	21.70704	0.97144	0.89202	0.90520
	60%	$1.0349 imes10^{-1}$	12.06025	0.78417	0.47157	0.48271
	Average	$1.9270  imes 10^{-2}$	26.84495	0.95753	0.88940	0.89405
	(100, 100) px	$5.4622\times 10^{-4}$	32.88275	0.99816	0.99555	0.99553
TRAN/ $(\Delta x, \Delta y)$	(250, 250) px	$5.4341 imes10^{-4}$	32.89609	0.99817	0.99558	0.99555
	(400, 400) px	$5.4083 imes10^{-4}$	32.93044	0.99818	0.99560	0.99557
	(550, 550) px	$5.3478\times10^{-4}$	32.99304	0.99820	0.99566	0.99562
	Average	$\textbf{5.4131} \times \textbf{10}^{-4}$	32.92558	0.99818	0.99560	0.99557

On the other hand, Figures 15–17 show some results of the watermarked and attacked images and the corresponding extracted watermark. The first two columns in each figure correspond to some satisfying examples of extracting the watermark, whereas the last two columns represent those cases where the extracted watermark presented worse performance. Moreover, each row corresponds to a different attack.



**Figure 15.** Results examples of the watermarked and attacked images (GF, MF, GN, SP) and the corresponding extracted watermark, showing satisfying results (**b**,**f**,**j**,**n**) and those with worse performance (**d**,**h**,**l**,**p**).

In addition, Tables 8–10, show the metrics obtained of the extracted watermarks for each attack of Figures 15, 16 and 17, respectively, both the most satisfying and worst performance result. It is essential to mention that the extracted watermarks are visually recognizable in most cases in the worse-performance results (the last two columns).



**Figure 16.** Results examples of the watermarked and attacked images (CE, EQ, JPEGC, SC) and the corresponding extracted watermark, showing satisfying results (**b**,**f**,**j**,**n**) and those with worse performance (**d**,**h**,**l**,**p**).

Attack/Parameter	Row (Figure 15)	Parameter Value	PSNR (dB)	NCC	MSSIM
$GF/(N \times N)$	1 (left)	$3 \times 3$	34.44387	0.99879	0.99697
	1 (right)	$9 \times 9$	16.77039	0.93424	0.77541
$MF/(N \times N)$	2 (left)	$3 \times 3$	34.40309	0.99878	0.99694
	2 (right)	$5 \times 5$	27.85757	0.99449	0.98804
$GN/(\sigma)$	3 (left)	0.1	34.37611	0.99877	0.99692
	3 (right)	0.5	21.46712	0.97659	0.92703
SP/(d)	4 (left)	0.6	34.26986	0.99874	0.99685
	4 (right)	0.75	16.62360	0.93240	0.72361

**Table 8.** Metrics of the extracted watermarks for attacked images from Figure 15.



**Figure 17.** Results examples of the watermarked and attacked images (CE, EQ, JPEGC, SC) and the corresponding extracted watermark, showing satisfying results (**b**,**f**,**j**) and those with worse performance (**d**,**h**,**l**).

Attack/Parameter	Row (Figure 16)	Parameter Value	PSNR (dB)	NCC	MSSIM
$CE/(P_{CE})$	1 (left) 1 (right)	4% $4%$	34.49884 12.99699	0.99881 0.85736	0.99701 0.51012
EQ/(LV)	2 (left)	128	34.40309	0.99878	0.99694
	2 (right)	8	28.03774	0.99472	0.98875
JPEGC/(QP)	3 (left)	90%	34.44387	0.99879	0.99697
	3 (right)	50%	13.45755	0.86972	0.69389
SC/(SF)	4 (left)	1.50x	34.40309	0.99878	0.99694
	4 (right)	0.25x	0.86910	0.03747	0.14016

 Table 9. Metrics of the extracted watermarks for attacked images from Figure 16.

Table 10. Metrics of the extracted watermarks for attacked images from Figure 17.

Attack/Parameter	Row (Figure 17)	Parameter Value	PSNR (dB)	NCC	MSSIM
$ROT/(\phi)$	1 (left)	$45^{\circ}$	34.45754	0.99880	0.99698
	1 (right)	$15^{\circ}$	9.70583	0.73919	0.30531
$CROP/(P_{CR})$	2 (left)	10%	34.55452	0.99882	0.99705
	2 (right)	50%	24.33087	0.98773	0.97136
TRAN/ $(\Delta x, \Delta y)$	3 (left)	(550, 550) px	34.52659	0.99882	0.99703
	3 (right)	(100, 100) px	29.46397	0.99619	0.99115

# 5.6. Comparison with Other Watermarking Methods

An important point to define the effectiveness of our proposal is to compare it with other similar algorithms, focusing on medical images. As we present in Section 5, our proposal was tested using fundus images, so this is an element to consider to selecting the approaches to compare. In Table 11, we present the comparison values of different watermarking schemes [9,14,16–19,22]. We present the results indicated in each paper using fundus images. In Table 12 we indicated the type of watermarking and the dataset repository that each algorithm uses.

 Table 11. Comparison values of different algorithms.

Watermarking Technique	PSNR (dB)	NCC	SSIM	MSE
Anushikha Singh et al. [9]	158.4183	1.0000	-	-
Zhen Dai et al. [10]	46.9631	-	-	-
A. George Klington et al. [14]	54.1572	-	-	-
Ranjana Dwivedi et al. [16]	46.8600	-	0.9914	-
Divyanshu Awasthi et al. [17]	39.4581 67.1475	0.9957 1.0000	0.9986 1.0000	$5.0534 \times 10^{-5}$ $8.2033 \times 10^{-8}$
Xiyao Liu et al. [18]	41.2995	0.9607	-	-
Muhammad Fachri et al. [19]	50.5228	0.9607	-	-
Payal Garg et al. [22]	51.7040	-	-	-
Proposed scheme	53.8638	0.9993	0.9937	$4.6976 \times 10^{-6}$

Table 12. Comparison of different algorithms about dataset and type of watermarking.

Technique	Image Dataset	Watermark Type	Capacity (Size)
Anushikha Singh et al. [9]	42	Digital patient ID	-
Zhen Dai et al. [10]	40	Binary image	$32 \times 32$
A. George Klington et al. [14]	1000	Fundus image & textual information	329,960 bits
Ranjana Dwivedi et al. [16]	10	Color image	$512 \times 512$
Divyanshu Awasthi et al. [17]	1	QR code	$256\times256$
Xiyao Liu et al. [18]	40	Hospital logo	$32 \times 32$
Muhammad Fachri et al. [19]	-	Binary image	64  imes 64
Payal Garg et al. [22]	1000	Fingerprints and gait images	-
Proposed scheme	1748	Binary image	$256\times256$

It is important to remark on some points of values reported by each method. In the method [14], the authors evaluated the three channels RGB, because the green channel is used for generating the watermark, and the other channels (blue and red) are used to insert the watermark. So in Table 11, we included the average value of red and blue channels. The paper [16] presents its results using different types of images, but the authors report their values metrics separately, so we only include the values referring to fundus images. The authors of the algorithm [17] used the LWT and tested their proposal using sub-bands LL and sub-bands HH, which is the reason why in Table 11 we present both results. The first row corresponds to sub-band LL and the second row is the value of sub-band HH. The methods [10,18] were evaluated with CT images, MRI images, ultrasound images, X-ray images, and fundus images and their results are the average of all of them, in these cases, we present this average. The authors of [10] indicated that used a total of 200 medical images, but we reported only the fundus images (40) that they used, in Table 12.

According to Tables 11 and 12, we can see, on the one hand, the authors do not calculate all metrics to evaluate their algorithm, so if we consider only PSNR values, even though the algorithms [9,14,17] have better PSNR values, our technique has a good result too. Also, considering the NCC value our technique is competitive because these values demonstrate that the image does not suffer visual modifications. In addition, the MSE is very low and SSIM is very close to one, so with these values, we reinforce the imperceptibility of the

watermark. On the other hand, to probe the effectiveness of an algorithm, it is necessary to use a considerable amount of data, in this case, our dataset is bigger than the others, and the capacity to hide information as a watermark is competitive compared to other state-of-the-art works. Finally, is important to compare taking into account the robustness of each one.

Particularly, the method described in [9] was not evaluated to probe its robustness, so this method has the highest value of PSNR (Table 11) but this does not facilitate guaranteed success against attacks. This paper only evaluates if the watermark affects the original diagnosis. According to its results, all tests are satisfactory indicating that the diagnosis does not change with the watermark.

Regarding the method reported by A. George Klington et al. [14], they only applied the jittering attack. They concluded that the proposed algorithm changes 43% of the original watermark bits against this attack. In the case of the method presented in [10], the metric they used to compare with other algorithms is mean BER (Bit Error Rate).

With the rest of the methods it is not easy to make a comparison because the authors do not use the same attacks or in some cases the parameters of each attack are different. In Table 13, Table 14 and Table 15 we include the results after applying the Median Filter attack, JPEG Compression, and Cropping attack, respectively.

Table 13. Comparison of different algorithms with our algorithm using Median Filter.

Technique	Parameter Value	NCC
Ranjana Dwivedi et al. [16]	$2 \times 2 \\ 3 \times 3$	0.9937 0.8945
Xiyao Liu et al. [18]	$3 \times 3 \\ 5 \times 5$	0.9759 0.9215
Proposed scheme	$2 \times 2$ $3 \times 3$ $5 \times 5$	0.9980 0.9981 0.9980

Table 14. Comparison of different algorithms with our algorithm using JPEG Compression.

Technique	Parameter Value	NCC
Ranjana Dwivedi et al. [16]	90%	0.9988
Xiyao Liu et al. [18]	80% 70%	0.9982 0.9939
Proposed scheme	90% 80% 70%	0.9971 0.9450 0.8331

Table 15. Comparison between Xiyao Liu et al. [18] and our algorithm after Cropping attack.

Technique	Parameter Value	NCC
Xiyao Liu et al. [18]	10% 20%	1.0000 0.9803
Proposed scheme	10% 20%	0.9982 0.9982

As we can see from Table 13, in all cases, our algorithm has the best results after applying the Median Filter. Concerning JPEGC, our results are low compared with other proposals ([16,18]), however, the metric values demonstrate that it is possible to recover the watermark and it is clear, as we demonstrate in Section 5. In the case of the Cropping attack, the method presented by Xiyao Liu et al. [18] has a better performance than our

algorithm, the values indicated that the extracted watermark is very similar to the original watermark. So with this experiment, we demonstrated that our method is competitive with other similar methods. Regarding the attacks such as Gaussian Noise, and Salt and Pepper Noise, the parameters that we use to test our method are more strict as we can see in Table 5. While methods [16,17] use Salt and Pepper densities of 0.05, 0.01, and 0.02. For Gaussian Noise, they use values of 0.01 and 0.02 or 0.03. So, according to our results, our method is robust against attacks.

# 5.7. Analysis of the Influence of the Watermark on the Diagnosis of Diabetic Retinopathy

In this section, we evaluate the influence of the watermarking process on DR diagnosis using a deep learning approach. Thus, accuracy is reported as a support metric, as the main focus remains on the effects of the watermarking on the inference produced by the different models.

## 5.7.1. Yolov8 Model Results

In the evaluation of YoloV8, it was observed that the model's top 1 prediction accuracy averaged 0.736. This is visually represented in the accompanying figure (Figure 18).



Figure 18. YOLOv8 top 1 accuracy.

Notably, the inference performance of this model was exemplary, accurately classifying the majority of data into their respective classes as delineated description files of the dataset, which specifies the class to which each image belongs. This high level of accuracy is further evidenced by the model's ability to generate consistent inference matrices for both marked and non-marked images, as illustrated in Figure 19.



Figure 19. YoloV8 (a) Original image. (b) Watermarked image. (c) Inference match.

The inference results for an original image (Figure 19a) and its watermarked counterpart (Figure 19b) are particularly revealing. The original and the watermarked images were classified as class 3 (severe). Both classifications align accurately with the DR grades assigned in the dataset labels (see Figure 19c). Crucially, the similarity in the inference matrices for both the marked and non-marked images demonstrates that watermarking does not adversely affect the model's inference capabilities on this architecture.

Therefore, the YoloV8 model exhibits high accuracy in its top 1 predictions and maintains consistent performance in classifying both watermarked and non-watermarked images, thereby proving that watermarking does not affect the predicted class.

## 5.7.2. VGG16 Model Results

The VGG16 model achieved a peak accuracy of 58.31% as we demonstrated in the results. However, this accuracy level did not show any improvement after the initial two epochs. This stagnation in performance is clearly illustrated in Figure 20, where the graphical representations of the model's accuracy and loss are presented.



Figure 20. VGGC16 inference results. (a) VGG16 Accuracy. (b) VGG16 Loss.

Further analysis of the VGG16 model was conducted to assess the impact of image watermarking on the classification performance. The comparison involved using both watermarked and non-watermarked images in the inference process. The results, as depicted in Figure 21, indicate that watermarking does not adversely affect the model's ability to correctly classify images for computer-aided DR diagnosis. Both watermarked and non-watermarked images yielded identical confidence levels in the predicted class, in the example class 0 (NO DR).



Figure 21. VGG16 inference results. (a) Original image. (b) Watermarked image. (c) Inference match.

# 5.7.3. InceptionV3 Model Results

In Figure 22, we present the performance metrics for the InceptionV3 model. This figure is divided into two parts: Figure 22a illustrates the accuracy of the InceptionV3 model, and Figure 22b displays its loss metrics. Both sub-figures provide a comprehensive view of the model's performance characteristics.



Figure 22. InceptionV3 inference results. (a) InceptionV3 Accuracy. (b) InceptionV3 Loss.

The InceptionV3 inference results indicate a slight but notable variation in the confidence level of the predicted class for certain images, with a deviation of  $\Delta E = \pm 1.9941\%$ . Despite this fluctuation, the classification accuracy remains high, as evidenced in Figure 23. It is important to note that while there is a minor difference in confidence values for certain images, the model consistently produces high-confidence results. This variation does not significantly impact the overall accuracy, as demonstrated in the comparative analysis of marked and non-marked images.



Figure 23. InceptionV3 inference results. (a) Original image. (b) Watermarked image. (c) Inference mismatch.

Further analysis, as shown in Figure 23, confirms the aforementioned slight discrepancy in the confidence levels between marked and non-marked images. While this variance is minimal and does not affect the correct classification of images, it is an interesting observation that non-marked images tend to exhibit marginally higher confidence levels. This pattern is not consistent across all images, as some show identical inference results for both their marked and non-marked versions as the one presented in Figure 24.



Figure 24. InceptionV3 inference results. (a) Original image. (b) Watermarked image. (c) Inference match.

## 5.7.4. ResNet50 Model Results

The ResNet50 model, trained over 100 epochs, achieved an outstanding accuracy of 1.0. An analysis of the inference matrix for both marked and non-marked images revealed a minuscule average difference ( $\Delta E = \pm 0.002\%$ ), indicating a negligible impact of watermarking on the model's performance. Notably, ResNet50 demonstrated exceptional

confidence in its predictions, averaging a 99.2% confidence level for the top predicted class. Remarkably, some images were classified with a perfect confidence score of 100%. The performance metrics for the ResNet50 model, including accuracy and loss, are comprehensively detailed in Figure 25.



Figure 25. ResNet50 inference results. (a) ResNet50 Accuracy. (b) ResNet50 Loss.

Figures 26 and 27 further elucidate the effect of watermarking on the inference process. In these figures, we observe the minimal impact of watermarking on the inference matrix, as well as instances where watermarking has no perceivable effect on the model's inference accuracy.



Original - Mild DR [0.9999435]

Marked - Mild DR [0.9999615]

(c)

Figure 26. ResNet50 slight inference results. (a) Original image. (b) Watermarked image. (c) Inference match.



**Figure 27.** ResNet50 complete inference results. (**a**) Original image. (**b**) Watermarked image. (**c**) Inference match.

In the evaluation of the ResNet50 model, the model achieved an unmatched accuracy rate of 100% over 100 epochs. A critical aspect of our analysis was examining the impact of watermarking on image classification. Remarkably, the ResNet50 model demonstrated remarkable resilience to watermarking, with an average deviation of just 0.002% in the inference matrix between marked and non-marked images. This negligible impact underscores the model's ability to maintain accuracy and reliability in classification, despite minor alterations in the input data. Furthermore, the model consistently displayed high confidence in its predictions, with an average confidence level of 99.2% for the top predicted class, and in some cases, reaching a perfect confidence score of 100%. The minimal

variation in performance, irrespective of watermarking, affirms that watermarking medical images using the proposed method does not present an effect in computer-aided and vision-based diagnostics.

## 6. Discussion

The method described in this paper focuses on a method to watermark digital medical images (fundus images) and four trained deep-learning models to evaluate if the original diagnosis is modified by the watermark. To probe the performance of the proposed method, typical metrics employed to evaluate quantitatively (MSE, PSNR and NCC) and visually (SSIM and MSSIM) any alteration of the watermarked image and the extracted watermark were used. Thus, the results reported in Table 3 show that the average performance metrics, using the 1748 fundus images, present high quality of the watermarked images with values of the MSE =  $4.6976 \times 10^{-6}$  PSNR = 53.8638 dB, NCC = 0.9993, SSIM = 0.9937, and MSSIM = 0.9938, demonstrating that the invisibility of the watermarks on the whole image dataset. On the other hand, the recovered watermarks in all marked images had excellent performance reporting values of MSE =  $7 \times 10^{-4}$ , PSNR = 32.0690 dB, NCC = 0.9975, SSIM = 0.9937, and MSSIM = 0.9943. This can be verified, in Figures 13 and 14, where are given some examples of the best and worst performing, respectively. Thus, the reporting PSNR values, for these watermarked images, were between 53 and 54 dB for both the best and worst results. Regarding the extracted watermarks, the PSNR values were between 33 and 34 dB for the best results, and they decreased between 21 and 30 dB for the worst ones, but with extracted watermarks visually clear and discernible.

To evaluate the robustness of the proposed watermarking algorithm, the more common attacks in this application were applied. Thus, both image processing and geometric attacks were employed by varying their parameter from low and high effect over the watermarked images. Among the image processing attacks were applied: Gaussian and Median Filter, Gaussian and Salt and Pepper Noise, Contrast Enhancement, Histogram Equalization, JPEG Compression, and Image Scaling. Additionally, Rotation, Cropping, and Translation operations were applied as geometric attacks. Thus, from Tables 5–7, the average metrics show that the watermark is successfully extracted for the Median Filter, Gaussian and Salt and Pepper Noise, Histogram Equalization, Rotation, and Translation attacks, obtaining PSNR average values between 31.40636 and 32.92558 dB as a quantitative metric and MSSIM average values between 0.98996 and 0.99557 as a qualitative metric, allowing to extract the watermark with high visual quality as it is shown in Figures 15–17. For the Contrast Enhancement attack, average values of PSNR = 29.29871 dB and MSSIM = 0.91870 (edge values) were obtained (see Table 6 and Figure 16). In addition, for the Image Cropping attack (Table 7 and Figure 17), average values of PSNR = 26.84495 dB and MSSIM = 0.89405 are consistent because up to 60% of the image is removed, destroying the watermark as a consequence.

On the other hand, worse performance was obtained for the following attacks: Gaussian Filter (PSNR = 28.24112 dB, MSSIM = 0.90795) and JPEG Compression (PSNR = 22.30630 dB, MSSIM = 0.78557). However, in both cases it is possible to recover the watermark and distinguish the visual information in it for parameters with high effect, as shown in Figures 15d and 16l, respectively. In the Gaussian Filter case, the low values are caused by a  $9 \times 9$  window when the edges of the images are considerably affected. For a JPEG Compression attack, the low values are present for quality percentages less than 70% due to information loss.

Finally, for Image Scaling, Table 6 shows the lowest average values of PSNR = 19.56431 dB and MSSIM = 0.66882, which is because the watermark extraction is unsuccessful for scaling factor values less than 1 or image downsampling (see Figure 16p), affecting the general performance for this attack. Succinctly, in image downsampling, it is not possible to recover the watermark. This has a similar behavior to the Cropping attack where the information of the image is also removed.

To determine if our algorithm is competitive with other similar approaches, we compare it with other methods taking into consideration their use of fundus images, because an important key of our investigation is to determine if the watermark modifies the original diagnosis. So as a first step, we compare the different methods [9,10,14,16-19,22] with our proposal to evaluate the invisibility of the watermarks. The results in Tables 11 and 12 demonstrate that our technique is competitive with the state-of-the-art, including the contribution that our algorithm was tested to determine if the watermark modifies the original diagnosis. Regarding robustness, we presented the outcomes of the algorithms employing identical parameters to our proposal, see Tables 10, 14 and 15. The values in these attacks indicated that our algorithm is robust and it is possible to extract the watermark without visible modifications, having the worst value when we applied a JPEG Compression of 70%. In the case of the Median Filter, we have the best results. With the rest of the attacks that we select to test our algorithm, the parameters employed are more strict than the parameters used by other proposals, so it demonstrates the robustness of our method. Referring to the capacity of the watermark, we use a logo and its dimensions are similar to the watermarks employed in the state-of-the-art.

Regarding the effect of watermarking in diagnosing diabetic retinopathy, we used various deep-learning models by evaluating their inference capabilities rather than solely concentrating on accuracy metrics.

Thus, the YoloV8 model displayed noteworthy performance, with a top 1 prediction accuracy averaging 0.736. This high accuracy level was maintained consistently across both watermarked and non-watermarked images. The model adeptly classified the majority of the data into their respective classes, as defined in the dataset description file. An in-depth examination of the inference results revealed that watermarking did not detrimentally affect the model's ability to accurately classify images, a critical aspect in medical image diagnostics.

On the other hand, the VGG16 model showed a peak accuracy of 58.31%, a figure that notably stagnated after the initial two epochs. This stagnation is visually depicted in the corresponding figures. Despite this, the VGG16 model effectively classified both watermarked and non-watermarked images, with no significant difference in the confidence levels of the predicted classes. This finding was vital in understanding the impact of watermarking in medical image analysis, especially in the context of diabetic retinopathy.

The performance of the InceptionV3 model was also evaluated. Although there was a slight variation in confidence levels between watermarked and non-watermarked images, this deviation was minor and did not significantly influence the model's overall classification accuracy. The InceptionV3 model successfully managed to maintain high-confidence results across different image types, thus underscoring its robustness in handling watermarked medical images.

Lastly, the ResNet50 model demonstrated exceptional performance, achieving a 100% accuracy rate over 100 epochs. The analysis of the inference matrix for both marked and non-watermarked images revealed a minuscule average difference, indicating the minimal impact of watermarking on the model's performance. The model displayed extraordinary confidence in its predictions, further affirming its reliability in accurately classifying medical images irrespective of watermarking.

Additionally, this finding is encouraging for the future of automated medical diagnosis systems. It suggests that these systems can handle watermarked images without losing accuracy, which is important for maintaining patient confidence and data protection.

## 7. Conclusions and Future Work

This study introduces a novel, bio-inspired watermarking algorithm for fundus images leveraging the transform domain. Our approach integrates a combination of mathematical tools to achieve robust and imperceptible watermarking without compromising clinical diagnosis. Firstly, the Steered Hermite transform acts as a natural detector for blood vessels, enabling automatic insertion of the watermark within this region of interest. This ensures imperceptibility while maintaining robustness. We further enhance security by incorporating the Jigsaw transform, an encryption process that scrambles the watermark information.

Secondly, Singular Value Decomposition bolsters the watermark's resilience against various image processing and geometric attacks. Our algorithm's effectiveness was evaluated on the MESSIDOR-2 dataset, encompassing 1748 RGB fundus images. Results confirmed superior invisibility and robustness compared to existing state-of-the-art methods.

Critically, we investigated the impact of watermarking on DR diagnosis. Four deep learning models (YoloV8, VGG16, InceptionV3, and ResNet50) were trained and tested on watermarked and original images. Remarkably, the analysis revealed a negligible influence on both disease classification accuracy and confidence scores. This finding underscores the compatibility of digital watermarking with computer-aided diagnostic systems, paving the way for enhanced data security without compromising diagnostic quality.

Looking forward, we aim to explore methods for strengthening the watermark's resistance against attacks that manipulate or remove image information, potentially impacting the watermark itself. Additionally, we will delve deeper into blood vessel-based watermarking to optimize both imperceptibility and robustness.

By integrating these advancements, we can contribute to a future where robust and secure medical image management coexists seamlessly with accurate clinical diagnosis, ultimately benefiting patient care and overall healthcare security.

Author Contributions: Conceptualization, E.M.-A., S.L.G.-C. and J.B.; methodology, E.M.-A., S.L.G.-C., J.B. and A.L.-F.; software, E.M.-A., S.L.G.-C., J.B. and A.L.-F.; validation, E.M.-A., S.L.G.-C., J.B. and A.L.-F.; formal analysis, E.M.-A., S.L.G.-C. and J.B.; investigation, E.M.-A., S.L.G.-C., J.B. and A.L.-F.; resources, E.M.-A., S.L.G.-C. and A.L.-F.; data curation, E.M.-A., S.L.G.-C. and A.L.-F.; writing—original draft preparation, E.M.-A., S.L.G.-C., J.B. and A.L.-F.; writing—original draft preparation, E.M.-A., S.L.G.-C., J.B. and A.L.-F.; writing—original draft preparation, E.M.-A., S.L.G.-C., J.B. and A.L.-F.; writing—review and editing, E.M.-A., S.L.G.-C., J.B. and A.L.-F.; usualization, E.M.-A., S.L.G.-C. and A.L.-F.; supervision, E.M.-A. and S.L.G.-C.; project administration, E.M.-A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research has been funded by Universidad Panamericana under the Program "Fomento a la Investigación UP 2021" grant UP-CI-2021-MEX-24-ING.

**Data Availability Statement:** The image dataset used in this work is publicly available from the Kaggle public dataset MESSIDOR-2: https://www.kaggle.com/datasets/geracollante/messidor2/ (accessed on 10 August 2023).

Acknowledgments: Ernesto Moya-Albor, Jorge Brieva, and Alberto Lopez-Figueroa would like to thank Facultad de Ingeniería of Universidad Panamericana for all support in this work. Sandra L. Gomez-Coronel thanks Instituto Politécnico Nacional (UPIITA) for the support in this work. The image dataset was kindly provided by the MESSIDOR program partners (see https://www.adcis.net/en/third-party/messidor/, accessed on 10 August 2023).

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

#### Abbreviations

The following abbreviations are used in this manuscript: ACNN Atrous Convolutional Neural Network

- AMD Age-Related Macular Degeneration
- ANN Artificial Neural Network
- BER Bit Error Rate
- BRB Blood-Retina Barrier
- CAD Computer-Aided Diagnosis
- CE Contrast Enhancement
- CLAHE Enhance Local Contrast
- CNN Convolutional Neural Network
- CROP Cropping

DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DME	Diabetic Macular Edema
DR	Diabetic Retinopathy
DWT	Discrete Wavelet Transform
EPR	Electronic Patients Record
EQ	Histogram Equalization
FDCuT	Fast Discrete Curvelet Transforms
GF	Gaussian Filter
GN	Gaussian Noise
HT	Hermite Transform
HVS	Human Vision System
ICDR	International Clinical Diabetic Retinopathy
IHT	Inverse Hermite Transform
ISHT	Inverse Steered Hermite Transform
JPEGC	JPEG Compression
JST	Jigsaw Transform
LBP	Local Binary Patterns
LTP	Local Ternary Patterns
LV	Equalization Levels number
LWT	Lifting Wavelet Transform
M-CNN	Multipath Convolutional Neural Network
MF	Median Filter
MSE	Mean Square Error
MSSIM	Mean Structural Similarity Index
NCC	Normalized Cross-Correlation
PCE	Percent Saturation
PCR	Cropping Percentage
PSNR	Peak Signal to Noise Ratio
QP	Quality Percentage
RDM	Recursive Dither Modulation
RDWT	Redundant Discrete Wavelet Transform
RNN	Recurrent Neural Network
ROI	Region Of Interest
RONI	Region Of No Interest
ROT	Rotation
SC	Image Scaling
SDLTP	Spherical Directional Local Ternary Pattern
SF	Scaling Factor
SHT	Steered Hermite Transform
SLT	Slantlet Transform
SP	Salt and Pepper Noise
SSIM	Structural Similarity Index
SVD	Singular Value Decomposition
SVM	Support Vector Machine
TRAN	Translation
WHO	World Health Organization
YOLO	You Only Look Once

## References

- World Health Organization. *Blindness and Vision Impairment*; Technical report; World Health Organization: Geneva, Switzerland, 2021. Available online: https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment (accessed on 10 August 2023).
- 2. Harvard Health Publishing. *Retinopathy;* Technical report; Harvard Medical School: Boston, MA, USA, 2020. Available online: https://www.health.harvard.edu/a\_to\_z/retinopathy-a-to-z (accessed on 10 August 2023).
- 3. The Pan American Health Organization. *Prevention of Blindness and Eye Care—Blindness*; Technical report; The Pan American Health Organization: Washington, DC, USA, 2019. Available online: https://www.paho.org/hq/index.php?option=com\_content& view=article&id=13693:prevention-blindness-eye-care-blindness&Itemid=39604&lang=en (accessed on 10 August 2023).

- 4. Salamanca, O.; Geary, A.; Suárez, N.; Benavent, S.; Gonzalez, M. Implementation of a diabetic retinopathy referral network, Peru. *Bull. World Health Organ.* **2018**, *96*, 674–681. [CrossRef]
- Carrillo-Alarcón, L.; López-López, E.; Hernández-Aguilar, C.; Martínez-Cervantes, J. Prevalencia de retinopatía diabética en pacientes con diabetes mellitus tipo 2 en Hidalgo, México. *Rev. Mex. Oftalmol.* 2011, 3, 125–178.
- Prado-Serrano, A.; Guido-Jiménez, M.; Camas-Benítez, J. Prevalencia de retinopatía diabética en población mexicana. *Rev. Mex.* Oftalmol. 2009, 83, 261–266.
- 7. Hervella, A.S.; Rouco, J.; Novo, J.; Ortega, M. Learning the retinal anatomy from scarce annotated data using self-supervised multimodal reconstruction. *Appl. Soft Comput.* **2020**, *91*, 106210. [CrossRef]
- 8. Mookiah, M.R.K.; Acharya, U.R.; Chua, C.K.; Lim, C.M.; Ng, E.; Laude, A. Computer-aided diagnosis of diabetic retinopathy: A review. *Comput. Biol. Med.* 2013, 43, 2136–2155. [CrossRef]
- 9. Singh, A.; Dutta, M.K. Imperceptible watermarking for security of fundus images in teleophthalmology applications and computer-aided diagnosis of retina diseases. *Int. J. Med. Inform.* **2017**, *108*, 110–124. [CrossRef]
- Dai, Z.; Lian, C.; He, Z.; Jiang, H.; Wang, Y. A Novel Hybrid Reversible-Zero Watermarking Scheme to Protect Medical Image. IEEE Access 2022, 10, 58005–58016. [CrossRef]
- 11. Liu, J.; Li, J.; Ma, J.; Sadiq, N.; Bhatti, U.A.; Ai, Y. A Robust Multi-Watermarking Algorithm for Medical Images Based on DTCWT-DCT and Henon Map. *Appl. Sci.* **2019**, *9*, 700. [CrossRef]
- 12. Dey, N.; Ahmed, S.S.; Chakraborty, S.; Maji, P.; Das, A.; Chaudhuri, S.S. Effect of trigonometric functions-based watermarking on blood vessel extraction: An application in ophthalmology imaging. *Int. J. Embed. Syst.* **2017**, *9*, 90–100. [CrossRef]
- Nayak, J.; Subbanna Bhat, P.; Acharya, U.R.; Sathish Kumar, M. Efficient storage and transmission of digital fundus images with patient information using reversible watermarking technique and error control codes. *J. Med. Syst.* 2009, 33, 163–171. [CrossRef] [PubMed]
- 14. Klington, G.; Ramesh, K.; Kadry, S. Cost-Effective watermarking scheme for authentication of digital fundus images in healthcare data management. *Inf. Technol. Control* 2021, *50*, 645–655. [CrossRef]
- 15. Singh, A.; Dutta, M.K.; Sharma, D.K. Unique identification code for medical fundus images using blood vessel pattern for tele-ophthalmology applications. *Comput. Methods Programs Biomed.* **2016**, 135, 61–75. [CrossRef] [PubMed]
- Dwivedi, R.; Srivastava, V.K. An Imperceptible Semi-blind Color Image Watermarking Using RDWT and SVD. In Proceedings of the International Conference on VLSI, Communication and Signal Processing, Prayagraj, India, 14–16 October 2022; Springer: Prayagraj, India, 2022; pp. 283–293.
- Awasthi, D.; Srivastava, V.K. Robust and Imperceptible Color Image Watermarking Using LWT, Schur Decomposition, and SVD in YCbCr Color Space. In Proceedings of the International Conference on VLSI, Communication and Signal Processing, Prayagraj, India, 14–16 October 2022; Springer: Prayagraj, India, 2022; pp. 259–271.
- Liu, X.; Lou, J.; Fang, H.; Chen, Y.; Ouyang, P.; Wang, Y.; Zou, B.; Wang, L. A Novel Robust Reversible Watermarking Scheme for Protecting Authenticity and Integrity of Medical Images. *IEEE Access* 2019, 7, 76580–76598. [CrossRef]
- Mahyudin, M.F.; Novamizanti, L.; Sa'idah, S. Robust Watermarking using Arnold and Hybrid Transform in Medical Images. In Proceedings of the 2021 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT), Bandung, Indonesia, 27–28 July 2021; pp. 180–185. [CrossRef]
- 20. Decencière, E.; Zhang, X.; Cazuguel, G.; Laÿ, B.; Cochener, B.; Trone, C.; Gain, P.; Ordóñez Varela, J.R.; Massin, P.; Erginay, A.; et al. Feedback on a publicly distributed image database: The Messidor database. *Image Anal. Stereol.* **2014**, *33*, 231–234. [CrossRef]
- Abràmoff, M.D.; Folk, J.C.; Han, D.P.; Walker, J.D.; Williams, D.F.; Russell, S.R.; Massin, P.; Cochener, B.; Gain, P.; Tang, L.; et al. Automated analysis of retinal images for detection of referable diabetic retinopathy. *JAMA Ophthalmol.* 2013, 131, 351–357. [CrossRef]
- 22. Garg, P.; Jain, A. A robust technique for biometric image authentication using invisible watermarking. *Multimed. Tools Appl.* **2023**, *82*, 2237–2253. [CrossRef]
- 23. Kapoor, P.; Arora, S. Applications of Deep Learning in Diabetic Retinopathy Detection and Classification: A Critical Review. *Lect. Notes Data Eng. Commun. Technol.* **2022**, *91*, 505–535. [CrossRef]
- 24. Kaggle. Diabetic Retinopathy Detection Competition. 2015. Available online: https://www.kaggle.com/c/diabetic-retinopathy-detection (accessed on 15 August 2023).
- Radha; Suchetha; Raman, R.; Madhumitha; Meena, S.; Sruthi; Philip, N. Classification of Retinal Lesions in Fundus Images Using Atrous Convolutional Neural Network. In *Futuristic Communication and Network Technologies*. VICFCNT 2020; Lecture Notes in Electrical Engineering; Springer: Singapore, 2022; Volume 792, pp. 551–564. [CrossRef]
- 26. Dutta, A.; Agarwal, P.; Mittal, A.; Khandelwal, S. Detecting grades of diabetic retinopathy by extraction of retinal lesions using digital fundus images. *Res. Biomed. Eng.* **2021**, *37*, 641–656. [CrossRef]
- 27. Gayathri, S.; Gopi, V.; Palanisamy, P. Diabetic retinopathy classification based on multipath CNN and machine learning classifiers. *Phys. Eng. Sci. Med.* **2021**, *44*, 639–653. [CrossRef]
- Chetoui, M.; Akhloufi, M. Explainable end-to-end deep learning for diabetic retinopathy detection across multiple datasets. J. Med. Imaging 2020, 7. [CrossRef]
- Wei, Q.; Li, X.; Yu, W.; Zhang, X.; Zhang, Y.; Hu, B.; Mo, B.; Gong, D.; Chen, N.; Ding, D.; et al. Learn to Segment Retinal Lesions and Beyond. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 7403–7410. [CrossRef]

- Ni, J.; Chen, Q.; Liu, C.; Wang, H.; Cao, Y.; Liu, B. An Effective CNN Approach for Diabetic Retinopathy Stage Classification with Dual Inputs and Selective Data Sampling. In Proceedings of the 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA), Boca Raton, FL, USA, 16–19 December 2019; pp. 1578–1584. [CrossRef]
- Randive, S.; Senapati, R.; Bhosle, N. Spherical Directional Feature Extraction with Artificial Neural Networkfor Diabetic Retinopathy Classification. In Proceedings of the 2018 IEEE 13th International Conference on Industrial and Information Systems (ICIIS), Rupnagar, India, 1–2 December 2018; pp. 152–157. [CrossRef]
- 32. Loheswaran, K. Optimized KFCM Segmentation and RNN Based Classification System for Diabetic Retinopathy Detection. In *ICCCE 2020*; Lecture Notes in Electrical Engineering; Springer: Singapore, 2021; Volume 698, pp. 1309–1322. [CrossRef]
- 33. Shorfuzzaman, M.; Hossain, M.; El Saddik, A. An explainable deep learning ensemble model for robust diagnosis of diabetic retinopathy grading. *ACM Trans. Multimed. Comput. Commun. Appl.* **2021**, *17*, 113. [CrossRef]
- 34. Suresh, M.; Indira, S.; Ramachandran, S. Classification of Fundus Images Based on Non-binary Patterns for the Automated Screening of Retinal Lesions. *Lect. Notes Netw. Syst.* **2021**, 204, 773–787. [CrossRef]
- 35. Sharif, M.; Shah, J. Automatic screening of retinal lesions for grading diabetic retinopathy. *Int. Arab. J. Inf. Technol.* 2019, 16, 766–774.
- Wang, R.; Chen, B.; Meng, D.; Wang, L. Weakly Supervised Lesion Detection From Fundus Images. *IEEE Trans. Med. Imaging* 2019, 38, 1501–1512. [CrossRef] [PubMed]
- Kaur, J.; Mittal, D. Estimation of severity level of non-proliferative diabetic retinopathy for clinical aid. *Biocybern. Biomed. Eng.* 2018, 38, 708–732. [CrossRef]
- DelaPava, M.; Ríos, H.; Rodríguez, F.; Perdomo, O.; González, F. A deep learning model for classification of diabetic retinopathy in eye fundus images based on retinal lesion detection. In *Proceedings of the 17th International Symposium on Medical Information Processing and Analysis, Campinas, Brazil, 17–19 November 2021*; SPIE: Bellingham, WA, USA, 2021; Volume 12088. [CrossRef]
- 39. Abdelmaksoud, E.; El-Sappagh, S.; Barakat, S.; Abuhmed, T.; Elmogy, M. Automatic Diabetic Retinopathy Grading System Based on Detecting Multiple Retinal Lesions. *IEEE Access* 2021, *9*, 15939–15960. [CrossRef]
- Biswas, S.; Upadhya, R.; Das, N.; Das, D.; Chakraborty, M.; Purkayastha, B. An Intelligent System for Diagnosis of Diabetic Retinopathy. *Adv. Intell. Syst. Comput.* 2020, 1139, 97–110. [CrossRef]
- Krause, J.; Gulshan, V.; Rahimy, E.; Karth, P.; Widner, K.; Corrado, G.S.; Peng, L.; Webster, D.R. Grader Variability and the Importance of Reference Standards for Evaluating Machine Learning Models for Diabetic Retinopathy. *Ophthalmology* 2018, 125, 1264–1272. [CrossRef] [PubMed]
- 42. Martens, J.B. The Hermite Transform-Theory. IEEE Trans. Acoust. Speech Signal Process. 1990, 38, 1595–1606. [CrossRef]
- Mira, C.; Moya-Albor, E.; Escalante-Ramírez, B.; Olveres, J.; Brieva, J.; Vallejo, E. 3D Hermite transform optical flow estimation in left ventricle CT sequences. *Sensors* 2020, 20, 595. [CrossRef]
- Hennelly, B.M.; Sheridan, J.T. Image encryption techniques based on the fractional Fourier transform. In *Optical Information Systems*; Javidi, B., Psaltis, D., Eds.; International Society for Optics and Photonics, SPIE: Bellingham, WA, USA, 2003; Volume 5202, pp. 76–87. [CrossRef]
- Chang, C.C.; Hu, Y.S.; Lin, C.C. A digital watermarking scheme based on singular value decomposition. In *Combinatorics, Algorithms, Probabilistic and Experimental Methodologies. ESCAPE 2007;* Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer: Berlin/Heidelberg, Germany, 2007; Volume 4614 LNCS, pp. 82–93. [CrossRef]
- Ovalle-Magallanes, E.; Avina-Cervantes, J.G.; Cruz-Aceves, I.; Ruiz-Pinales, J. Transfer learning for stenosis detection in X-ray Coronary Angiography. *Mathematics* 2020, 8, 1510. [CrossRef]
- Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. ImageNet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255. [CrossRef]
- Yosinski, J.; Clune, J.; Bengio, Y.; Lipson, H. How transferable are features in deep neural networks? In Advances in Neural Information Processing Systems 27 (NIPS '14); Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2014; pp. 3320–3328.
- Tajbakhsh, N.; Shin, J.Y.; Gurudu, S.R.; Hurst, R.T.; Kendall, C.B.; Gotway, M.B.; Liang, J. Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? *IEEE Trans. Med. Imaging* 2016, 35, 1299–1312. [CrossRef] [PubMed]
- Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* (*IJCV*) 2015, 115, 211–252. [CrossRef]
- 51. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. arXiv 2016, arXiv:1612.08242.
- Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015—Conference Track Proceedings, San Diego, CA, USA, 7–9 May 2015.
- Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9. [CrossRef]
- 54. Jocher, G.; Chaurasia, A.; Qiu, J. Ultralytics YOLOv8. 2023. Available online: https://github.com/ultralytics/ultralytics (accessed on 31 August 2023).

- 55. Gomez-Coronel, S.L.; Moya-Albor, E.; Pérez-Daniel, K.R.; Brieva, J.; Cruz-Aceves, I.; Hernandez-Aguirre, A.; Soto-Alvarez, J.A. Authentication of medical images through a hybrid watermarking method based on Hermite-Jigsaw-SVD. In Proceedings of the 18th International Symposium on Medical Information Processing and Analysis, Valparaíso, Chile, 9–11 November 2022; Brieva, J., Guevara, P., Lepore, N., Linguraru, M.G., Rittner, L., Castro, E.R., Eds.; International Society for Optics and Photonics, SPIE: St Bellingham, WA, USA, 2023; Volume 12567, p. 125671G. [CrossRef]
- 56. Poynton, C.A. A Technical Introduction to Digital Video; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 1996.
- 57. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. 2015. Available online: www.tensorflow.org (accessed on 31 August 2023).
- 58. Chollet, F. Keras. 2015. Available online: https://keras.io (accessed on 1 August 2023).
- 59. Horé, A.; Ziou, D. Image Quality Metrics: PSNR vs. SSIM. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 2366–2369. [CrossRef]
- 60. Rao, Y.R.; Prathapani, N.; Nagabhooshanam, E. Application of normalized cross correlation to image registration. *Int. J. Res. Eng. Technol.* **2014**, *3*, 12–16.

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