Article

Ensemble Learning-Based Solutions: An Approach for Evaluating Multiple Features in the Context of H&E Histological Images

Jaqueline J. Tenguam [1,*], Leonardo H. da Costa Longo [1], Guilherme F. Roberto [2], Thaína A. A. Tosta [3], Paulo R. de Faria [4], Adriano M. Loyola [5], Sérgio V. Cardoso [5], Adriano B. Silva [6], Marcelo Z. do Nascimento [6] and Leandro A. Neves [1,*]

1 Department of Computer Science and Statistics (DCCE), São Paulo State University (UNESP), Rua Cristóvão Colombo, 2265, São José do Rio Preto 15054-000, São Paulo, Brazil
2 Department of Informatics Engineering, Faculty of Engineering, University of Porto, Dr. Roberto Frias, sn, 4200-465 Porto, Portugal; guilhermefroberto@gmail.com
3 Science and Technology Institute, Federal University of São Paulo (UNIFESP), Avenida Cesare Mansueto Giulio Lattes, 1201, São José dos Campos 12247-014, São Paulo, Brazil; tosta.thaina@gmail.com
4 Department of Histology and Morphology, Institute of Biomedical Science, Federal University of Uberlândia (UFU), Av. Amazonas, S/N, Uberlândia 38405-320, Minas Gerais, Brazil; paulo.faria@ufu.br
5 Area of Oral Pathology, School of Dentistry, Federal University of Uberlândia (UFU), R. Ceará—Umuarama, Uberlândia 38402-018, Minas Gerais, Brazil
6 Faculty of Computer Science (FACOM), Federal University of Uberlândia (UFU), Avenida João Naves de Ávila 2121, Bl.B, Uberlândia 38400-902, Minas Gerais, Brazil; adrianobs@gmail.com (A.B.S.); marcelo.zanchetta@gmail.com (M.Z.d.N.)
* Correspondence: jaqueline.tenguam@unesp.br (J.J.T.); leandro.neves@unesp.br (L.A.N.)

Abstract: In this paper, we propose an approach based on ensemble learning to classify histology tissues stained with hematoxylin and eosin. The proposal was applied to representative images of colorectal cancer, oral epithelial dysplasia, non-Hodgkin's lymphoma, and liver tissues (the classification of gender and age from liver tissue samples). The ensemble learning considered multiple combinations of techniques that are commonly used to develop computer-aided diagnosis methods in medical imaging. The feature extraction was defined with different descriptors, exploring the deep learning and handcrafted methods. The deep-learned features were obtained using five different convolutional neural network architectures. The handcrafted features were representatives of fractal techniques (multidimensional and multiscale approaches), Haralick descriptors, and local binary patterns. A two-stage feature selection process (ranking with metaheuristics) was defined to obtain the main combinations of descriptors and, consequently, techniques. Each combination was tested through a rigorous ensemble process, exploring heterogeneous classifiers, such as Random Forest, Support Vector Machine, K-Nearest Neighbors, Logistic Regression, and Naive Bayes. The ensemble learning presented here provided accuracy rates from 90.72% to 100.00% and offered relevant information about the combinations of techniques in multiple histological images and the main features present in the top-performing solutions, using smaller sets of descriptors (limited to a maximum of 53), which involved each ensemble process and solutions that have not yet been explored. The developed methodology, i.e., making the knowledge of each ensemble learning comprehensible to specialists, complements the main contributions of this study to supporting the development of computer-aided diagnosis systems for histological images.

Keywords: ensemble learning; handcrafted features; deep-learned features; two-stage feature selection method; histological images

1. Introduction

Histopathological analysis involves procedures that aim to investigate tissue samples that are commonly stained with specific dyes, such as hematoxylin and eosin (H&E) [1,2].

In these processes, specialists identify unusual alterations in the structures of cells and highlight potential abnormal health conditions, such as the diagnosis of cancer. It is noteworthy that this type of disease is a significant cause of early deaths worldwide and has high social and economic costs [3]. For instance, cancer is the second leading cause of death in the United States [4]. Thus, the early detection of diseases often enables less invasive treatments and increases the possibility of finding a cure and/or patient survival. The required steps in the preparation process of H&E images can influence the presentation of histological aspects, further increasing the difficulty of accurately diagnosing diseases under investigation. In addition to this problem, the analysis process takes time and may be susceptible to subjective interpretations by specialists [5–7]. These interpretation problems are mainly caused by human issues, such as subjectivity and fatigue. On the other hand, computer-aided diagnosis (CAD) methods play a fundamental role in this task since they can support specialists with second opinions [8–10], especially regarding H&E images [5,8,11–17].

In this regard, two categories of descriptors are typically investigated for the development of CAD systems. The first category consists of handcrafted features (HFs) defined by distinct extraction methods, usually aiming to overcome specific problems [18–22]. Among the HFs, it is possible to highlight techniques based on fractal geometry that use multiscale and multidimensional methods (Higuchi fractal dimension, probabilistic fractal dimension, box fusion fractal dimension, lacunarity and percolation) [23–29], Haralick [30] and local binary patterns (LBPs) [31]. For instance, Haralick and LBPs have been applied in several imaging contexts [32–34], exploring the identification of lung cancer subtypes [35], the presence of cancerous characteristics in breast tissue samples [18,36] and the classification of colorectal cancer [6]. In addition, techniques that involve fractals at multiple scales and/or dimensions have also been applied to quantify the pathological architectures of tumors [23,25,26], demonstrating relevant results in the pattern recognition of prostate cancer [37], lymphomas [38], intraepithelial neoplasia [39], breast tumors [40], colorectal cancer [13] and psoriatic lesions [41]. Moreover, fractal methods are important for texture analysis because they provide information about the complexity of textures and patterns that are similar at various levels of magnification [42].

The second group of descriptors consists of deep-learned (DL) features obtained using convolution neural networks (CNNs) [43]. This group has been useful for defining different CAD approaches [15,17,22,44–48] that consider data representations with multiple levels of abstraction [8,47,49]. The most explored models have provided the best accuracy rates on the ImageNet dataset [50], such as AlexNet [51], GoogleNet [52], VGGNet [53], ResNet [54], Inception [52] and other applications [20,32,55].

Despite the advances provided individually by DL features and HFs, investigations have been carried out to develop models based on combinations of these features [18,21,32,55–58], generating strategies known as ensemble learning (EL) [20,59]. Moreover, these studies have indicated no single solution for distinct datasets. EL models can also consider distinct classification algorithms in order to obtain more accurate single decisions by applying ensembles of classifiers [20,60]. This type of association has provided important results in the study of cervical cell imaging [20]. Another highlight is the model presented by [59], who conducted a comparative study between a logistic regression classifier trained only with DL features, an ensemble of HFs and an ensemble of all features. The authors concluded that the ensemble involving all features delivered the best distinction rates. In the context of using fractal techniques with CNN models, the method presented by [29] considered an ensemble involving two CNN architectures: one pre-trained with histological images and the other pre-trained with artificial images, which were generated using features from fractal techniques. The authors concluded that their proposal outperformed classification algorithms and CNN models when applied separately.

It is essential to observe that EL models can be developed with the most relevant features, exploring a single selection stage to reduce the search space and increase the accuracy of the system [13,15,18]. The use of feature selection through two stages can

also be implemented [61–63]. In these cases, the strategies were applied to the ensemble of DL, and the results obtained were better than those obtained via a single stage. Thus, this approach can provide a reduced number of main compositions for developing CAD systems, making knowledge even more comprehensive for specialists. Moreover, the use of this combination is not restricted to image analysis with promising solutions in the frequency domain [64].

The versatility of EL strategies was observed with different combinations between features and classifiers to investigate some types of histological images [20,21,29,32,55,56,59,65,66], but not as described here, which were addressed to define patterns of techniques through multiple H&E datasets. Some examples of EL strategies that can still be investigated are the HFs, DL features, and the HFs and DL features, all of which are in a classifier ensemble context. The best configuration can be compared via classifications only with the use of CNN models, which are useful to indicate the pertinence of using ensemble learning in pattern recognition of various H&E images (colorectal cancer, oral dysplasia, non-Hodgkin's lymphoma, and liver tissue). Moreover, the previously indicated EL models can also be explored via feature selection in two stages, which is a valuable approach to present more optimized solutions, in addition to significantly reducing a high-dimensional search space, such as those explored here. Thus, known problems such as overfitting or underfitting are minimized [67]. An EL model capable of providing the main solutions for various H&E images, with robust computational approaches for pattern recognition, can significantly improve CAD systems and make knowledge more comprehensive for specialists.

This work presents an EL approach to classify histological images from different contexts. The proposal explored multiple handcrafted features through multidimensional and multiscale fractal techniques (Higuchi fractal dimension, probabilistic fractal dimension, box fusion fractal dimension, lacunarity, and percolation), Haralick and LBPs descriptors, and deep-learned descriptors, which were obtained from several convolutional neural network architectures. Moreover, a two-stage feature selection (ranking with metaheuristic algorithms) with a heterogeneous ensemble of classifiers completed the proposed method to indicate the best solutions. The first stage of selection was defined through the ReliefF algorithm. In the second phase, the approach to discover the most effective features within each reduced subset was employed, exploring particle swarm optimization, genetic algorithm, and binary gray wolf optimization. Each result was verified through a robust ensemble process with a Support Vector Machine, Naive Bayes, Random Forest, Logistic Regression, and K-Nearest Neighbors. This proposal provided the following contributions:

1. An EL approach not yet explored in H&E image classification, able to identify the primary combinations of features via two-stage feature selection (ranking with metaheuristics) with a heterogeneous ensemble of classifiers;
2. Best ensembles of descriptors to distinguish multiple histological datasets that have been stained with H&E;
3. An analysis of the proposal's usefulness concerning relevant models available in the specialized literature with indications of the best performances concerning colorectal cancer, oral epithelial dysplasia, and gender classification from liver tissue. This was achieved by utilizing a limited number of features, ranging from 11 to 29 attributes;
4. A more robust baseline approach, with solutions without overfitting, which is useful in evaluating and composing new approaches for pattern recognition in histological images;
5. A breakdown of the main descriptors present in the best ensembles, making the knowledge comprehensive for specialists and helpful in improving CAD systems.

Section 2 presents the proposed methodology, providing information about the techniques used to compose the ensemble learning approach. Section 3 shows the results and engages in a discussion following the application of this approach. Finally, Section 4 indicates the main findings and suggestions for future exploration.

2. Methodology

In this proposal, the first stage considered techniques for feature extraction. The explored models to compose the handcrafted descriptors included multidimensional and multiscale fractal approaches as well as Haralick and LBPs. The DL features were collected via different CNN architectures with the transfer learning strategy. The explored CNN models were ResNet-50, VGGNet-19, DenseNet-121, Inception v3, and EfficientNet-B2. In the next stage, the descriptors were organized to define the ensembles of features, which were evaluated through a feature selection process based on two stages: ranking with metaheuristic approaches. Lastly, an ensemble of classifiers was determined to indicate the main EL solutions within the scope of various H&E datasets. Figure 1 summarizes the main stages of our proposal. Details are presented in the next sections.
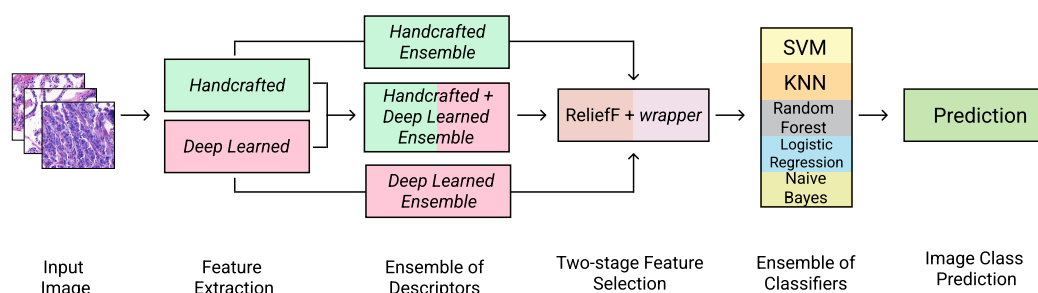

Figure 1. Illustrative overview of the proposed approach.

2.1. Handcrafted Fractal Features

Among the various fractal techniques found in the specialized literature, models from a multidimensional and multiscale perspective, such as fractal dimension (probabilistic [24,41], box-merging [68], and Higuchi [12,28,39]), lacunarity [24,41], and percolation [38] were considered in this proposal, since they allow a complementary quantification of color images. Some image types, such as histological medical images, are spectrally and spatially complex and often show certain similarities at different spatial scales. From the chosen approaches, fractal geometry allows the study and description of irregular or fragmented forms of elements in nature as well as complex objects that Euclidean geometry cannot analyze [26]. Moreover, the combinations of fractal features explored here are highly capable of quantifying histological information, such as those existing in the H&E datasets investigated in this study [13,29].

Fractal dimension is often applied to evaluate the irregularity and complexity of a region under analysis, enabling the quantification of the fractional filling of a structure in some scale interval. The lacunarity attribute quantifies the deviation of the translational invariance of a geometric object, indicating how similar the parts of different regions of the objects are to each other. Thus, images with low-value lacunarity are more homogeneous regarding size distribution and the spatial arrangement of gaps. They are also translation invariant, since the sizes of the holes are equal [24,41]. Finally, percolation theory is useful in characterizing many disordered systems, as the percolation process is purely random. Therefore, the locations of the system under analysis to be occupied or that will remain empty are randomly chosen with probability p, allowing the formation of clusters [69]. When a cluster presents a path connecting the two ends of the system, it is considered that there has been percolation. The obtained topology from this process has structures highly related to the fractals. These concepts are also applied in image analysis, considering percolation through connectivity among neighboring pixels [38,70].

2.1.1. Probabilistic Fractal Dimension

The fractal dimension based on a probabilistic approach (DF_p) was determined following the method outlined in [24,41], which involves the gliding box process. In this technique, for a given color image provided as input, utilizing the RGB color model, each

pixel was represented by a 5D vector (x, y, r, g, b) with spatial coordinates (x, y) corresponding to the color components (r, g, b). Using a hypercube with side $L = 3$, the image was scanned in steps of one pixel from the top left to the bottom right. After this, the size of the hypercube was incremented by two units until it reached L_{max}. An illustration of this process is presented in Figure 2.

Figure 2. A grid map with overlapping boxes (red), $L = 3$ in (**a**) and $L = 5$ in (**b**), in order to illustrate the gliding box process.

Each pixel i of a hypercube, $F_i = f(x_i, y_i, r_i, g_i, b_i)$, was compared to the corresponding central pixel $F_c = f(x_c, y_c, r_c, g_c, b_c)$ using a distance measure. Here, the chessboard (Δ_{ches}), Euclidean (Δ_{eucl}) and Manhattan (Δ_{manh}) distances were considered, according to Equations (1)–(3), respectively. Each *pixel* F_i with a distance Δ less than or equal to L was labeled as *1* to indicate that it belongs to the hypercube under analysis. Otherwise, the assigned label was *0*. By counting the pixels labeled as *1*, it was possible to define a matrix $P(m, L)$, which characterized the probability P that m points belong to a hypercube with side L. The result was a structure according to the illustration available in Table 1.

$$\Delta_{ches} = max(|F_i(o_i) - F_c(o_c)|), \ o \in r, g, b, \tag{1}$$

$$\Delta_{eucl} = \sqrt{\sum_o (F_i(o_i) - F_c(o_c))^2}, \ o \in r, g, b, \tag{2}$$

$$\Delta_{manh} = \sum_o |F_i(o_i) - F_c(o_c)|, \ o \in r, g, b. \tag{3}$$

Thus, the matrix $P(m, L)$ was normalized according to Equation (4), ensuring that the sum of the elements of a column is equal to *1*.

$$\sum_{m=1}^{L^2} P(m, L) = 1, \forall L. \tag{4}$$

From the matrix $P(m, L)$, it was possible to obtain the local fractal dimension $N(L)$ from each size L through Equation (5). This quantification allowed obtaining the angular coefficient of the linear regression, defined by $log \ L \times log \ N(L)$, as the probabilistic fractal dimension DF_p of the image under analysis.

$$N(L) = \sum_{m=1}^{L^2} \frac{P(m, L)}{m}. \tag{5}$$

Table 1. Illustration of a probability matrix P.

	3	**5**	**...**	$\boldsymbol{L_{max}}$
1	$P(1,3)$	$P(1,5)$	...	$P(1,L_{max})$
2	$P(2,3)$	$P(2,5)$	...	$P(2,L_{max})$
⋮	⋮	⋮	⋱	⋮
L^2	$P(L^2,3)$	$P(L^2,5)$	⋮	$P(L^2,L_{max})$

2.1.2. Box-Merging Fractal Dimension

The fractal dimension with box fusion approach (DF_n) was applied as described by [68] to quantify a color image within the RGB color space. Each image axis was divided into s partitions to establish a partition table. For instance, the partitions on the x-axis occurred according to Equation (6):

$$t_x = \lfloor \frac{x}{\epsilon_x} \rfloor = \lfloor \frac{xs}{L_x} \rfloor, \tag{6}$$

where t_x indicates the x axis partition, x is the coordinate of any pixel in the box, L is the observation scale, and ϵ refers to the L/s ratio.

The resulting table considered the coordinates of all partitions with at least one element. Identical lines were grouped, resulting in a total of z distinct lines. Subsequently, the regression $log\, z \times log\, s$ was defined, and the corresponding angular coefficient indicated the value of DF_n for an image.

2.1.3. Multidimensional and Multiscale Higuchi Fractal Dimension

The fractal dimension, proposed by Higuchi (DF_H) in 1988 [71], enables the analysis of time series in the 1D domain. The method proved relevant results for biological signal analysis [27,72,73]. The approach explored here considered the approach presented in [28], which expanded the analysis process by including a multidimensional and multiscale strategy.

The applied model considered sets of 1D series from an image I, with dimensions h x w. The multiscale step was defined using different observation scales. The procedure involved sliding vectors to analyze each pixel series. The initial position of the sliding vector was determined from the first position of the pixel series under analysis, and it was incremented by one position to the right until its last position coincided with the last position in the series. The vector lengths were represented as l, such that $3 \leq l \leq min(h,w)$, given the need for each sliding vector to have a central pixel.

In this process, as previously indicated, each pixel of the series under analysis was represented as a 5D vector (x,y,r,g,b) [28,41], allowing an analysis between the pixels involved in each transition of the sliding vector by the series. This procedure characterized the multidimensional approach, which is a useful strategy to quantify the patterns in color images. In each iteration in the multidimensional approach, the central pixel of the sliding vector was compared to every other pixel inside it based on the Manhattan distance (Equation (3)). If Δ was less than or equal to the current sliding vector length, the central pixel composed an auxiliary subseries referring to the observation scale.

The Higuchi fractal dimension (DF_H) was defined through each finite auxiliary subseries extracted from the input series of an image I. In this approach, a finite series of discrete points can be defined as $X = x(1), x(2), x(3), \ldots, x(N_{DH})$, such that x indicates an element of the image I and N_{DH} indicates the maximum number of points available in the series under analysis. Thus, the method generated d new series X_{dj}, considering that the starting point occurred from j, and d is a granularity factor:

$$X_{dj} : x(j), x(j+d), x(j+2d), \ldots, x\left(j + \lfloor \frac{N_{DH} - j}{d} \rfloor \cdot d\right), \tag{7}$$

where $j = 1, 2, \ldots, d$.

For each X_{dj}, the curve length, $Len_j(d)$, was calculated such that:

$$Len_j(d) = \frac{1}{d} \left\{ \left(\sum_{i=1}^{\lfloor \frac{N_{DH}-j}{d} \rfloor} |x(j+id) - x(j+(i-1) \cdot d)| \right) \frac{N_{DH} - 1}{\lfloor \frac{N_{DH}-j}{d} \rfloor \cdot d} \right\}. \tag{8}$$

Thus, *Len* was defined as the length of the curve for the interval *d*, which was calculated as the average value across the *d* sets of $Len_j(d)$:

$$Len(d) = \frac{1}{d} \sum_{j=1}^{d} Len_j(d). \tag{9}$$

The DF_H value for a subseries was given by the angular coefficient obtained from the linear regression of a $\ln Len(d) \times \ln d$ plot, which was obtained through linear least squares fitting. The *t* auxiliary pixel subseries from each initial input set was employed to calculate the descriptor DF_H [28]. Consequently, there are *t* values of DF_H. The fractal dimension value of the input series, DF_{HSerie}, was defined as the average of these *t* values:

$$DF_{HSerie} = \frac{1}{t} \sum_{i=1}^{t} DF_{H_i}, \tag{10}$$

where *t* indicates the number of values that *l* can take.

The fractal dimension value for the color image was given by averaging the values of all DF_{HSerie} [28].

2.1.4. Lacunarity

The lacunarity (*Lac*) was based on the approach from [41,74], considering the same probability matrix indicated previously in Section 2.1.1 to represent the multidimensional and multiscale strategy. This descriptor was obtained through the first-order (Equation (11)) and second-order (Equation (12)) moments based on the distribution measure given by Equation (13).

$$\lambda(L) = \sum_{m=1}^{L^2} mP(m, L), \tag{11}$$

$$\lambda^2(L) = \sum_{m=1}^{L^2} m^2 P(m, L), \tag{12}$$

$$\Lambda(L) = \frac{\lambda^2(L) - (\lambda(L))^2}{(\lambda(L))^2}. \tag{13}$$

2.1.5. Multidimensional and Multiscale Percolation

The percolation attribute (*Perc*) was computed using the approach outlined in [38]. Percolation theory was employed to analyze connected pixel paths stretching from one edge to another within an image. The applied method explored a multiscale approach using the gliding box technique. Thus, hypercubes were initially defined with $L = 3$ [13,38] were incremented by two units after the complete scan of the image. The relation of the quantity *T* of hypercubes that traversed an image with height *H* and width *W*, as a function of *L*, was given by the following:

$$T(L) = (H - L + 1) \times (W - L + 1), L \leq min(H, W). \tag{14}$$

For each hypercube with size *L*, a multidimensional approach was applied considering the most relevant color channel, according to the RGB model, aiming to perform a comparison against the central pixel P_c, as presented in [24,41] and Section 2.1.1. The comparison was also defined from the three distances mentioned in Equations (1)–(3). Therefore, when

the distance Δ assumed a value less than or equal to L, the pixel P was labeled with -1, indicating that the pixel represents a pore.

The percolation clusters were obtained based on the labeling of the Hoshen-Kopelman algorithm, as described in [38]. Once this labeling was given, a cluster was defined by considering neighboring pixels with the same label. When the first cluster has been identified, the algorithm advances to the next unverified pore. From this process, three functions were extracted: the average cluster C; the percolating box ratio Q; and the average coverage ratio of the largest cluster M. The calculation of the average number of clusters per box $C(L)$ was obtained by the number of clusters in a single box (c_i), given a scale L, divided by the total number of boxes T, according to the equation:

$$C(L) = \frac{\sum_{i=1}^{T(L)} c_i}{T(L)}. \tag{15}$$

The value of the percolating box ratio Q was obtained by counting the number of percolating boxes for each value of L. A box q_i was counted to increment $Q(L)$ if the ratio between the number of pixels labeled as pores (Ω_i) and the total number of pixels (L^2) is greater than the percolation threshold p:

$$q_i = \begin{cases} 1, & \frac{\Omega_i}{L^2} \geq 0.59275, \\ 0, & \frac{\Omega_i}{L^2} < 0.59275. \end{cases} \tag{16}$$

The ratio of percolating boxes as a function of L ($Q(L)$) was obtained by dividing the total number of percolating boxes q_i by the total number of boxes T on a scale L:

$$Q(L) = \frac{\sum_{i=1}^{T(L)} q_i}{T(L)}. \tag{17}$$

Finally, the average coverage ratio of the largest cluster (M) was calculated by identifying the coverage ratio of the largest cluster in each box evaluated at the L scale, according to Equation (18), where γ_i indicates the largest cluster in a i box.

$$M(L) = \frac{\sum_{i=1}^{T(L)} \gamma_i/L^2}{T(L)}. \tag{18}$$

2.1.6. Metrics Obtained from the Descriptor Curves

Fractal descriptors based on the probabilistic approach, lacunarity, and percolation were calculated with L scale variations, according to the *gliding box* method. In these cases, the value $L_{max} = 41$ was considered in this investigation [29]. The multidimensional and multiscale Higuchi fractal dimension was determined with l scale variations of the sliding vector, also with a maximum value of 41, and a granularity factor $d = 8$ [28]. The values from each approach were used to define feature curves for each image under analysis. In the Higuchi fractal dimension and lacunarity approaches, the curves were composed of their respective local values concerning each size of the sliding vector/hypercube. Regarding the percolation approach, the obtained curves were C, Q, and M, referring to the percolating regions.

In addition, each curve was represented by scalar values to compose the descriptor vectors [38,75]. We used the following extracted metrics: area under the curve (A), skewness (S), area ratio (Γ), maximum point (MP), and maximum point scale (MPS). The area under the curve (A) was calculated by the trapezoidal numerical integration method given by Equation (19).

$$A = \int_a^b f(x)dx \approx \frac{b-a}{2N} \sum_{n=a}^{b-1} (f(x_n) + f(x_{n+1})), \tag{19}$$

where a and b denote the minimum and maximum L values, respectively, and N denotes the number of samples.

The skewness (S) metric defines the asymmetry of a sample concerning its mean value. A negative skewness corresponds to a sample where the more frequent values occur on the lower part of the interval range, and the opposite applies if the skewness is positive. A completely symmetric sample yields a skewness of 0. Given a sample with N values, skewness was defined by Equation (20), where $\overline{x}$ denotes the mean value of the sample, and x_i represents the i-th value of x.

$$S = \frac{\frac{1}{N} \sum_{i=a}^{b} (x_i - \overline{x})^3}{\sqrt{\left[\frac{1}{N-1} \sum_{i=a}^{b} (x_i - \overline{x})^2 \right]^3}}. \tag{20}$$

The area ratio (Γ) was calculated as the ratio of the area of the right half of the curve to the area of the left half of the curve. This was calculated by considering the area under the curve ($Area_{a,b}$) between two points a and b on the x axis, as represented by Equation (21).

$$Area = \int_{a}^{b} f(x) dx. \tag{21}$$

From this definition, the area ratio Γ was calculated according to Equation (22).

$$\Gamma = \frac{A_{(b/2+1, b)}}{A_{(a, b/2)}} \tag{22}$$

The two last measures were obtained from the maximum point (MP) and its respective observation scale L (MPS).

2.2. Haralick Features

Haralick descriptors were derived from the co-occurrence matrix [30]. This matrix allowed us to verify the transition between two pixels in gray levels, u and v, considering a distance measure δ and an arbitrary angle θ. It is important to highlight that in the histology area, there have been studies exploring the interpixel distance variation GLCM textural parameter approach, such as the work proposed by [76]. This study has shown that Haralick descriptors present pattern differences between the cell nuclei for one interpixel distance. Moreover, the one interpixel distance allows the definition of fine-grained texture and regularity and demands reduced computational effort, especially when dealing with large images. Thus, in the proposed framework, the parameter values of distance $\delta = 1$ and direction $\theta = 0°, \theta = 45°, \theta = 90°$ and $\theta = 135°$ [13,62] were applied. Considering the co-occurrence matrices and the widely known descriptors from [30], the features investigated here were: (1) Angular Second Moment; (2) Contrast; (3) Correlation; (4) Sum of Squares; (5) Inverse Difference Moment; (6) Sum Average; (7) Sum Variance; (8) Sum Entropy; (9) Entropy; (10) Difference Variance; (11) Difference Entropy; (12) and (13) are Information Measures of Correlation; and (14) Maximal Correlation Coefficient. For the final 14 descriptors, the averages of the descriptors obtained in the four matrices were considered.

2.3. Local Binary Patterns

LBPs is used to generate a set of features that represent how binary patterns are distributed in a circular neighborhood around the center pixel. The result is an LBP code for each pixel of the image [31,77]. The neighborhood is defined by a radius R and the number of neighbors PN [78]. If the intensity of the neighboring pixel is greater than or equal to the reference intensity, the binary value 1 is assigned. Otherwise, the value 0 is assigned. Concatenating all the values in the region under analysis, starting from the top left corner and moving clockwise, results in a binary number. The corresponding decimal value takes values between 0 and 255 and represents the analyzed region. Therefore, the LBPs strategy was applied as described in [78,79] with a radius value of

$R = 1$ and considering $PN = 8$ neighbors. From these parameters, it was possible to obtain a total of 59 descriptors for each of the images under analysis, according to the method "extractLBPFeatures(Image)", which is available in the Matlab tool [80].

2.4. Deep-Learned Features

One of the main problems CNN models face is the number of samples required in the training stage. In this work, this issue was resolved through transfer learning, which is a strategy that uses knowledge obtained in one or more source tasks to improve the learning process in a target task [81,82]. CNN architectures belong to the class of inductive learning algorithms, mapping input features between classes to obtain a generalization of the data. Thus, inductive learning can be transferred from an architecture trained on the source task to the target class by adjusting the model space and correcting the bias values. Usually, this procedure is performed by replacing the last layer of the model (classifier) [32], providing shorter training time and suppressing the possibility of overfitting due to a small number of samples. The different CNN models explored here followed this strategy.

Therefore, the deep-learned descriptors were computed using transfer learning [82], as it allowed analyses involving image sets with a reduced number of samples. Here, the deep-learned features were obtained from CNN architectures pre-trained on the ImageNet dataset [29,32]. Five distinct CNN architectures were considered: ResNet-50 [54], VGG-19 [53], Inception v3 [83], DenseNet-121 [84], and EfficientNet-B2 [85]. Additional information is summarized in Table 2. These networks have shown relevant results in medical image classification problems in varied contexts [32,55,56] and also in histological image classification [20,29,86].

For the ResNet-50 architecture, the deep-learned descriptors were obtained from the last convolutional layer (avgpool) at the point just before the fully connected layer, as explored in other works [32,59,86]. Regarding the VGG-19 architecture, the features resulted from the second fully connected layer (FC2) [87,88]. In the Inception v3 and DenseNet-121 architectures, the descriptors were extracted at the point just before the last fully connected layer [89,90], indicating the last avgpool and global avgpool, respectively. Finally, the deep-learned features explored in the EfficientNet-B2 architecture were extracted from the last fully connected layer (FC) [91].

Table 2. Summary of each CNN model investigated in this study.

CNN Model	Layers	Parameters [92]	Input Dimensions
ResNet-50	50	25×10^6	$224 \times 224 \times 3$
VGG-19	19	143×10^6	$224 \times 224 \times 3$
Inception v3	48	23×10^6	$299 \times 299 \times 3$
DenseNet-121	121	8×10^6	$224 \times 224 \times 3$
EfficientNet-B2	342	9.2×10^6	$260 \times 260 \times 3$

2.5. Ensemble of Descriptors

In this study, we explored combinations of handcrafted and deep-learned features through an ensemble of descriptors to broaden the strategies of [21,32,55,56,60,66] with new patterns of techniques in multiple H&E images.

The following frameworks were used: (i) ensemble of handcrafted descriptors with an ensemble of classifiers; (ii) ensemble of deep-learned descriptors with an ensemble of classifiers; (iii) ensemble of handcrafted and deep-learned descriptors with an ensemble of classifiers. Moreover, CNN architectures were directly employed to classify H&E images to validate the effectiveness of the suggested approach within the context of colorectal cancer, oral epithelial dysplasia, non-Hodgkin's lymphoma, and liver tissues.

The total number of descriptors that comprised the ensembles depended on each category explored here. The handcrafted descriptors tally 462 values, including 389 fractal descriptors distributed as follows: percolation (225), lacunarity (75), probabilistic fractal

dimension (63), Higuchi fractal dimension (25), and box-merging approach for fractal dimension (1). Additionally, there are 59 LBPs descriptors and 14 Haralick descriptors. For the deep-learned features, they amount to 10,624 values. This total depends on the layers used in the analyses. Here, as described in Section 2.4, the chosen layers and their corresponding number of descriptors are summarized in Table 3.

Table 3. Number of extracted descriptors and their respective layer in each architecture.

CNN Architecture	Number of Descriptors	Extracted Layer
ResNet-50	2048	avgpool
VGGNet-19	4096	FC2
Inception v3	2048	last avgpool
DenseNet-121	1024	global avgpool
EfficientNet-B2	1408	FC

2.6. Two-Stage Feature Selection and Ensemble of Classifiers

The ensembles of descriptors were evaluated through the application of a two-stage feature selection process. The aim is to reduce the dimensionality of the feature space and identify the most suitable descriptors within each subset [62,93,94]. Initially, our approach involves ranking the feature sets through the ReliefF algorithm [95] and then applying a threshold to decrease the number of potential matches. The threshold considered the 100 best-ranked descriptors in each ensemble [13]. Therefore, the second phase intended to identify the optimal matches within each narrowed subset, which is accomplished through wrapper selection and the exploration of various metaheuristics, namely: genetic algorithm (GA) inspired by genetic evolution, particle swarm optimization (PSO) inspired by particle swarm behavior, and binary gray wolf optimization (bGWO) inspired by the hunting strategy of gray wolves [96]. We used the K-nearest neighbors algorithm to assess the effectiveness of each of these feature selection methods. This proposal allowed the investigation of possible patterns resulting from the combination of techniques in multiple H&E datasets [61–63]. Due to the stochastic characteristic of the wrapper selection algorithms, the final performance of a combination was determined from the arithmetic average of 10 different runs of the selection step. Also, the most selected features were defined by their frequencies in the ten runs [97].

2.6.1. Ensemble Classification

The final step of the method focused on evaluating each ensemble of descriptors derived from the preceding step. This assessment was conducted using a strong ensemble of classifiers, exploring five algorithms from distinct categories (based on the function, probability, decision tree, and instance-based K-nearest neighbors). In this case, the ensemble of classifiers was built upon the following algorithms: Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest, Logistic Regression, and Naive Bayes. The approach allowed combining results from individual classifiers to avoid overfitting [29,59,60]. The classifiers were individually trained, and the ensemble was applied taking into account the sum rule [29,56,60]. Figure 3 illustrates the evaluation process with the ensemble approach. The blocks indicated by dashed lines indicate that only one of the internal blocks was used in the composition of the model.

The assessment process for each classification algorithm involved the use of the cross-validation k-fold method, with $k = 3$ [62]. In this approach, the instances were divided into k independent groups, with each group further divided into a training set and a test set. In each iteration of the training process for k, one fold was used for evaluation, while the remaining $k - 1$ folds served as training data. Instances within each group were classified, resulting in k performance measures, one for each fold. The ultimate performance was determined by calculating the arithmetic average of these partial performances. The metrics used to evaluate each solution were the area under the ROC curve (AUC) and accuracy [98].
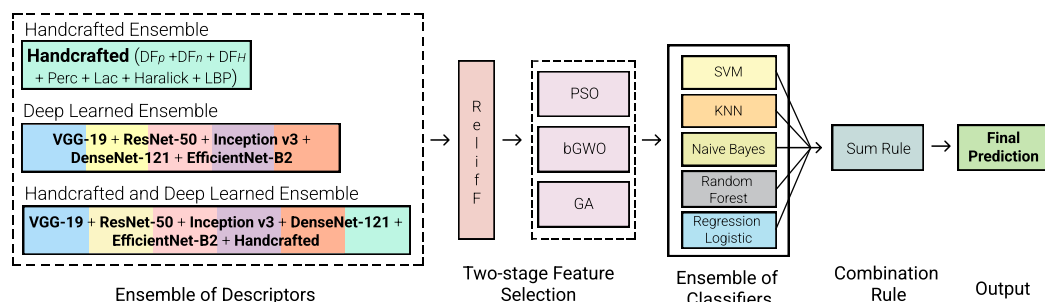
Figure 3. Diagram of the ensemble approach: boxes with dashed lines define that only one of its components was used to compose a solution.

2.7. Application Context

The proposed approach was employed to assess five histological image datasets, namely colorectal cancer (CR), oral epithelial dysplasia (OED), non-Hodgkin's lymphoma (NHL), and liver tissues, covering liver gender (LG) and liver aging (LA).

The CR dataset consists of histological images extracted from 16 sections stained with H&E, specifically focusing on T3 or T4 stage colorectal cancer. An expert in pathology labeled the images, categorizing them into benign or malignant groups. The dataset includes 165 images, with 74 images representing benign cases and 91 images representing malignant cases [99].

The OED dataset was created using 30 H&E histological sections from the tongues of mice that had been previously exposed to a carcinogen [100]. This study was approved by the Ethics Committee, protocol 038/39 of the Federal University of Uberlândia. A total of 148 ROIs divided into two classes, healthy (74) and severe dysplasia (74), were considered in this work.

The NHL dataset consists of 374 H&E images of malignant non-Hodgkin's lymphoma subdivided into three classes: Chronic Lymphocytic Leukemia (CLL) with 113 images; Follicular Lymphoma (FL) with 139 images; and Mantle Cell Lymphoma (MCL) with 122 images. The dataset is a collection of samples prepared by different histologists at various hospitals [101].

The last two datasets comprise liver tissues from 48 male and female mice obtained through the National Institute on Aging, Atlas of Gene Expression in Mouse Aging Project (AGEMAP). Images were acquired manually by a Carl Zeiss Axiovert 200 microscope at $40\times$ magnification. The liver gender (LG) dataset consists of 265 images, which were categorized into male (150) and female (115) classes of mice under an ad libitum diet. The liver aging (LA) dataset, on the other hand, contains 528 images of female mice at four ages: 1 (100), 6 (115), 16 (162) and 24 (151) months, also on an ad libitum diet [102].

An overview of the datasets is given in Table 4, indicating the names and their respective types, number of images, and number of classes. Some examples of images with their respective classes are presented in Figure 4.

Table 4. General information about the datasets used in this investigation.

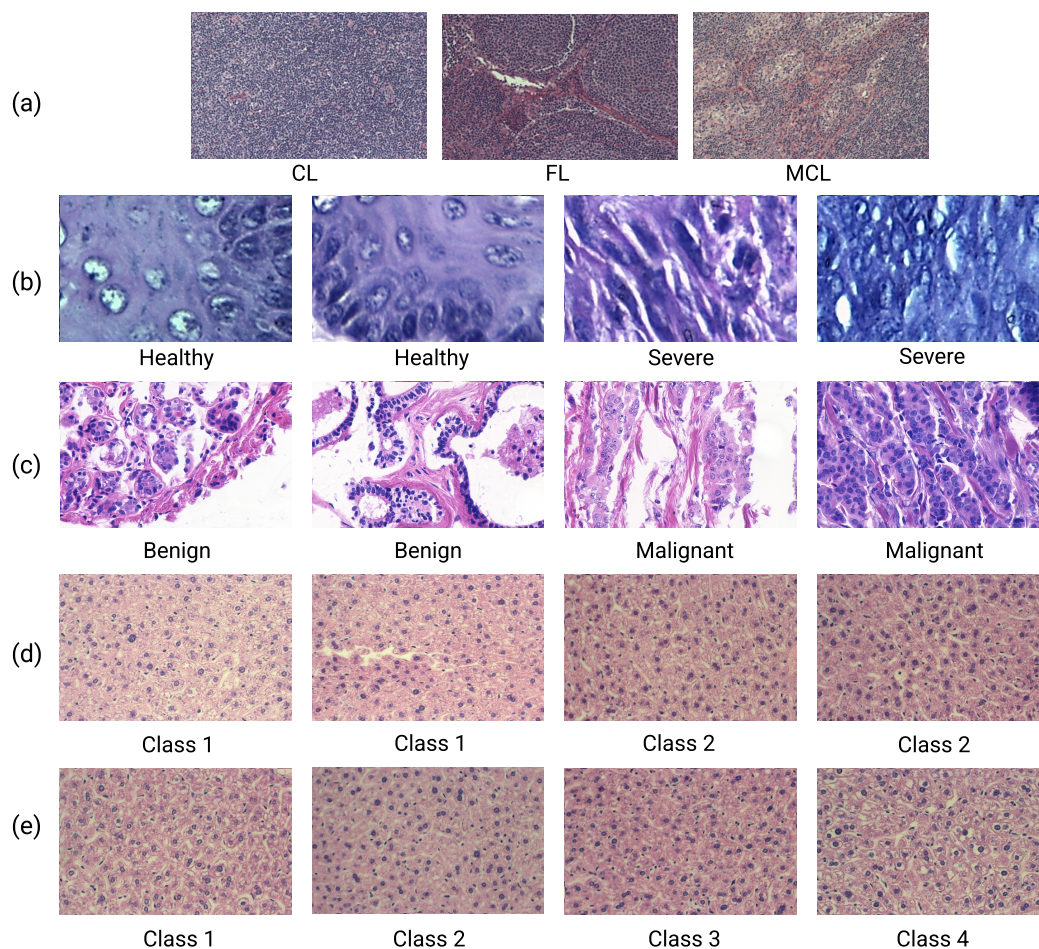| Dataset	Image	Number of Images	Number of Classes	Resolution
CR	Colorectal Tumor	165	2	567 × 430 to 775 × 522
OED	Oral Epithelial Dysplasia	148	2	450 × 250
LA	Liver Tissue	528	4	417 × 312
LG	Liver Tissue	265	2	417 × 312
NHL	Non-Hodgkin's Lymphoma	374	3	1388 × 1040

Figure 4. Examples of NHL [101] (**a**), OED [100] (**b**), CR [99] (**c**), LG (**d**) and LA (**e**) images [102].

2.8. Software Packages and Execution Environment

In this work, the handcrafted descriptors were implemented and extracted via MATLAB® R2019a [103]. The pre-trained convolutional neural network architectures were implemented via the Pytorch v1.10 library [104] in the Google Collaboratory execution environment [105]. Considering the CNN models, the package explored here uses a strategy in the first layer that allows the input of images of any size. This strategy avoided the loss of significant information from the images. Also, the feature selection and classification algorithms were employed using the Weka 3.8 [106] software. The default values suggested in each package were used except for specific ones mentioned in the text. The experiments were performed on an Intel Core i5 notebook, 8 GB RAM and 64-bit architecture operating system.

3. Results and Discussion

The ensemble learning approach was evaluated on five datasets of H&E histological images, as described in Section 2, with comparisons involving the different classes of each set. It is important to note that 45 types of tests were performed to explore different compositions of ensembles, including three associations of wrapper methods, in order to provide the main compositions among the 100 best-ranked descriptors with ReliefF. Each composition was evaluated via a heterogeneous ensemble of classifiers (Section 2.6.1). In Table 5, the average performances for each ensemble, considering the HFs and DL attributes, are shown. The best rates are highlighted in bold.

From Table 5, it is observed that the HFs and HFs+DL ensembles were responsible for the best results in four H&E datasets (OED, LA, LG, and NHL) out of the five investigated here. The accuracy values ranged from 90.72% to 100%. Thus, it is possible

to indicate that the handcrafted descriptors explored here (via HFs ensemble) are relevant for the classification process whether used separately or in combination with DL. The HFs ensemble provided the highest distinction rate on the OED dataset regardless of the wrapper selector taken as reference, indicating the optimal match on this dataset, which represents a further contribution of the proposed approach. On the other hand, HFs presented the lowest accuracy values (approximately 78%) in the NHL dataset, with three classes (CLL×FL×MCL), demonstrating a possible limit of HFs. When this category was combined with DL (HFs+DL ensemble), the result was the most expressive for the NHL dataset with an accuracy of 90.72%. Even so, this result represents an accuracy of at least 8% lower than the values achieved on other datasets. This is an important indication of the difficulties in distinguishing CLL×FL×MCL groups, especially considering only the HFs ensemble. Finally, when the DL ensemble is considered, the best solution was achieved in a single dataset (CR) but with an expressive rate (99.76% accuracy), illustrating its importance for the development of strategies to support the diagnosis of colorectal cancer. In addition, on this dataset, it is worth mentioning the HFs+DL ensemble as another potential solution, which achieved an accuracy of 99.58%, which was very close to that provided by the DL combination. This configuration represents an acceptable and common solution for different types of histological samples.

Table 5. Average accuracy values obtained through the proposed approach with two-stage feature selection (ReliefF + Wrapper).

Ensemble	Wrapper	CR	OED	LA	LG	NHL
	PSO	91.94%	**100%**	98.30%	99.36%	78.02%
HFs	GA	91.39%	99.66%	98.58%	99.36%	77.59%
	bGWO	92.30%	99.66%	**98.73%**	99.36%	77.94%
	PSO	99.45%	96.89%	93.50%	98.75%	90.64%
DL	GA	99.15%	97.64%	93.01%	98.75%	90.56%
	bGWO	**99.76%**	96.96%	93.39%	98.87%	90.13%
	PSO	99.58%	98.45%	96.86%	99.40%	90.51%
HFs+DL	GA	99.45%	98.45%	96.72%	99.36%	**90.72%**
	bGWO	99.45%	99.05%	96.44%	**99.47%**	90.43%

The best accuracy rates for each dataset are highlighted in bold.

To summarize the results discussed here, the best combinations of descriptors and selection algorithms are presented in Table 6, including the total number of descriptors, AUC, and accuracy averages. The accuracy values of the top 1 and top 10 solutions are also indicated, making it possible to observe the existing variation for the first and tenth solutions in each dataset, since the averages were calculated from the 10 best-ranked compositions in each dataset. It is important to emphasize that this ranking indicated the highest accuracy with the lowest number of descriptors.

Table 6. Main combinations in each H&E-stained histological dataset based on the criterion of higher accuracy with the lowest number of descriptors.

Dataset	Ensemble	Feature Selection	⌈Average of Descriptors⌉	Average Accuracy	Average AUC	Top 1 Accuracy	Top 10 Accuracy
CR	DL	ReliefF + bGWO	21 ± 3	$99.76\% \pm 0.30$	1	100%	99.39%
OED	HFs	ReliefF + PSO	11 ± 3	100%	1	100%	100%
LA	HFs	ReliefF + bGWO	40 ± 7	$98.73\% \pm 0.36$	$0.999 \pm 0.4 \times 10^{-3}$	99.24%	98.11%
LG	HFs+DL	ReliefF + bGWO	29 ± 6	$99.47\% \pm 0.30$	1	100%	98.87%
NHL	HFs+DL	ReliefF + GA	53 ± 5	$90.72\% \pm 1.04$	$0.980 \pm 1.5 \times 10^{-3}$	92.25%	89.57%

Based on the previously stated criterion, it can be reiterated that the HFs and HFs+DL ensembles were responsible for the best results in four of the five H&E-stained histological datasets investigated here. In these cases, the ReliefF + bGWO selection processes stand out with three occurrences. This indicates another pattern for the CR, LA, and LG

datasets, with expressive average AUC rates of 0.999 (LA) and 1 (CR and LG). In addition, the lowest top 10 accuracy was significant, with a rate above 98% (LA), on the dataset that involves four groups for the classification process. Moreover, the main solutions for these three datasets indicated a reduced number of descriptors, an average of 21 features for CR, 29 for LG, and 40 for LA.

Regarding the two-stage ReliefF + PSO, this strategy was the main solution for the OED dataset, providing maximum accuracy with the lowest number of descriptors among all solutions with only 11 features. When the two-stage ReliefF + GA is observed, this approach constitutes the best solution on a single dataset (NHL). In this case, the solutions explored 53 features on average, identifying the highest value among all solutions. The top 10 accuracy was 89.57%, and the top 1 was slightly better, 92.25%, reinforcing the difficulties present in this set. The NHL dataset comprises three classes (CLL×FL×MCL) and, possibly, with less heterogeneous histological patterns, implying more difficulties in the constitutions of the solutions. Even so, in this case, the average AUC was 0.98, which is an important value under the exposed conditions. For instance, considering that the original feature space had a range of 462 to 11,086 values, the outcome achieved in this study is another contribution that is capable of providing expressive average performances with few descriptors but highly relevant to the classification process.

3.1. Feature Occurrences in the Main Solutions: An Overview

To identify the descriptors present in the top solutions, as summarized in Table 6, a survey of the occurrence of each category of features in the first 10 solutions of each H&E-stained histological dataset was performed. To better understand the origin of each descriptor in each solution, the occurrences of the deep-learned descriptors are in Figure 5, CR, LG, and NHL datasets, and the handcrafted ones are in Figure 6, indicating solutions for LA, LG, OED, and NHL. It is important to highlight that the best solution for the LA dataset involved only HFs. Occurrences in NHL and LG also involved HFs due to the HFs+DL ensemble, justifying the representation of these datasets in Figure 6. Also, in these two datasets, the occurrence percentages were calculated relative to the total number of HFs in the HFs+DL ensemble, disregarding the percentages of deep-learned features.
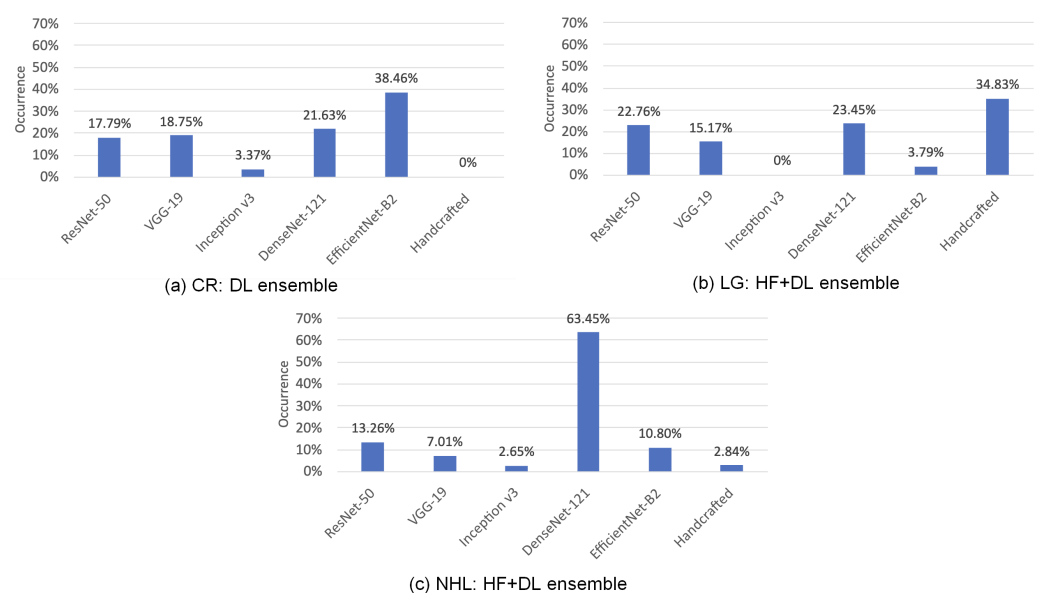

Figure 5. Occurrences of deep-learned features included in the best 10 solutions for the CR, LG, and NHL datasets.

Considering the distributions illustrated in Figure 5, it is possible to verify some behaviors. The lowest occurrences occurred concerning the Inception v3 network descriptors, a maximum of 3.37% for the CR dataset, and there were no instances among the best

10 solutions for the LG set. On the other hand, the descriptors via DenseNet-121 and EfficientNet-B2 networks have the highest occurrences, especially for the NHL dataset, in which 63.45% of the features originated from the DenseNet-121 model. Descriptors via the EfficientNet-B2 architecture stood out in the solutions for the CR dataset with 38.46% of the occurrences surpassing more homogeneous occurrences (from 17.79% to 21.63%) for the deep-learned features from the ResNet-50, VGG-19, and DenseNet-121. Another homogeneous distribution can be seen in the LG dataset, involving the same descriptor origins as the CR set. In this case, occurrences ranged from 15.17% to 23.45%. When DL versus HFs totalization is considered, it can be seen that DL attributes predominated in the solutions for the LG and NHL datasets with occurrences of 65.17% and 97.16%, respectively. Despite these differences, it is not possible to indicate that these were the most important in the classifications.
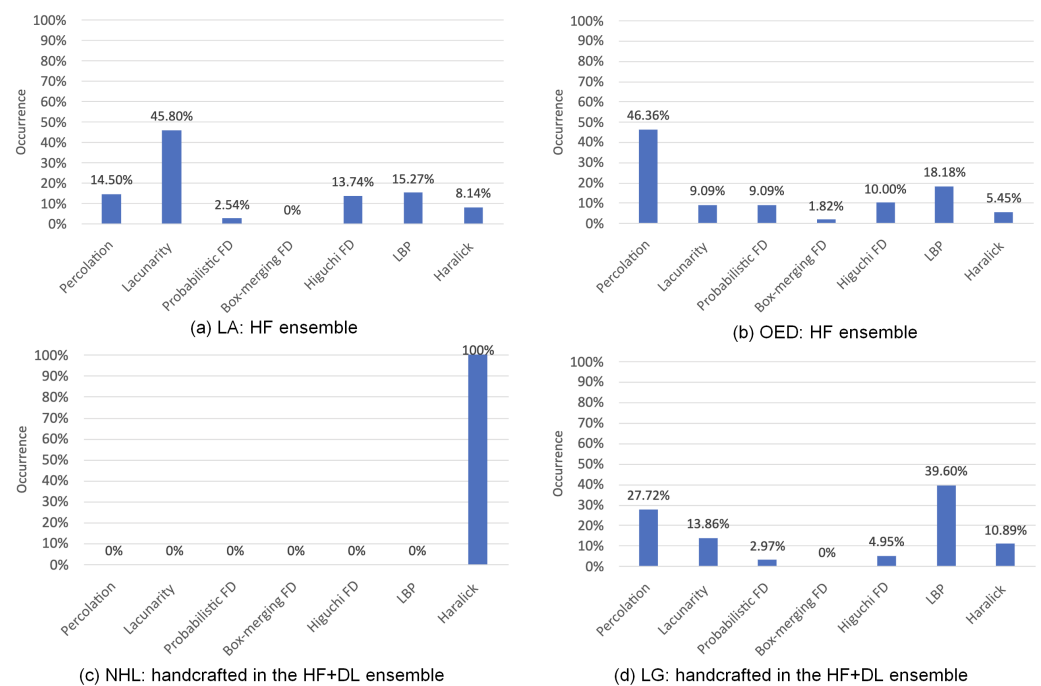

Figure 6. Occurrences of handcrafted features included in the best 10 solutions for the LA, OED, NHL, and LG datasets.

Concerning the occurrences of handcrafted descriptors (Figure 6), the lowest occurrence was of the box-merging fractal dimension (DF_n from [68]), since it was not selected for the top 10 solutions in three out of four histological datasets stained with H&E. This descriptor was present in the solutions for OED but with the lowest occurrence, only 1.84%. The probabilistic fractal dimension descriptor (DF_p from [24,41]) was the second lowest occurrence but constituted the solutions for three of the four histological datasets. Another interesting result involves the enhanced version of Higuchi fractal dimension descriptor (DF_H [28]) with occurrences that surpassed those of the DF_n and DF_p approaches, which are widely explored in the literature [24,41,68], contributing to advancements in this particular research field. Finally, it is possible to observe the descriptors with the highest occurrences for each H&E dataset: lacunarity (45.80%) for LA; percolation (46.36%) for OED; LBPs (39.60%) for LG; and with a highlight, Haralick as the only ones that constituted the solutions for NHL.

3.2. Performance Overview against Different Approaches

The best performances were observed respecting those obtained via traditional CNN architectures applied directly to the H&E-stained histological images. ResNet-50 [54], VGG-19 [53], Inception v3 [83], DenseNet-121 [84] and Efficient-Net-B2 [85] were the models

tested in this process, using the following: a fine-tuning process; cross-validation k-folds, with $k = 3$; 10 epochs; the stochastic gradient descent algorithm; initial learning rate of 0.01, decaying by 0.75 every two epochs; and, loss function as cross-entropy. Similar experiments were described by [29]. Also, the CNN models were applied using the transfer learning strategy [82], considering each network pre-trained on the ImageNet dataset [29,32]. Thus, each dataset with a reduced number of examples was investigated using each model after a fine-tuned process, mapping the last corresponding layer of each architecture with the groups available in each H&E dataset. The final connections with their weights were updated based on the total number of classes in each context, ensuring appropriate results without overfitting. In addition, the input images were normalized according to the mean and standard deviation values of the ImageNet dataset [107]. The accuracy values the networks provide are shown in Table 7.

Table 7. Accuracy provided by CNN architectures after fine-tuning process: best results are in bold.

CNN Architecture	CR	OED	LA	LG	NHL
ResNet-50	96.73%	**96.00%**	**91.48%**	97.78%	**74.27%**
VGG-19	**98.67%**	92.67%	77.96%	91.98%	65.07%
Inception v3	85.64%	90.50%	85.37%	92.59%	67.07%
DenseNet-121	97.45%	93.17%	87.41%	93.98%	72.40%
EfficientNet-B2	96.12%	94.67%	90.00%	**98.89%**	70.27%

The best accuracy rates for each dataset are highlighted in bold.

To understand the differences between the distinction rates of the models proposed here and those obtained via networks applied directly, we consider the average accuracy values achieved in each H&E-stained histological dataset, as summarized in Table 6. Thus, it is possible to verify that the accuracy values via the proposed approach overcome those provided by the ResNet-50, VGG-19, Inception v3, DenseNet-121, and EfficientNet-B2 networks. The classification rates with the convolutional networks ranged from 74.27% to 98.89%. Therefore, the gains in accuracy ranged from 0.6% to 16%, approximately. The smallest gain occurred in the CR dataset (0.58%) and the largest (16.45%) occurred in the NHL set. For example, we increased the classification rate in the NHL dataset, which involves three classes, from 74.27% to 90.72%, illustrating an additional contribution of this study.

Regarding the noted differences, the Friedman test was applied to verify if these solutions are statistically relevant. The Friedman test is a non-parametric statistical method capable of ranking the solutions under investigation, where the best option is set at the first position [108]. This type of test allows us to observe the variance of repeated measures and to analyze whether the existing differences are statistically significant via p-values. The smaller the p-value, the greater evidence that the difference is statistically relevant. It is possible to indicate that there is some relevant difference when the p-value is less than 0.05. In the experiments carried out here, the resulting p-value was 0.0004, indicating that the differences between the solutions are statistically significant.

In addition, the Friedman test ranks the solutions as a table. The result involving the experiments is displayed in Table 8, with Friedman's score indicated. The solutions obtained in this study are the most relevant for each dataset. It is important to note that when applying the Friedman test, each dataset represents a different sample (row) in relation to the corresponding solution. Each performance obtained through a solution in an H&E set has a rank value assigned based on the order of the best solutions. In the case of a tie, average ranks were assigned to the solutions. In each column, Friedman's score was calculated as the average of the ranks of the samples, providing a final score for each solution [108].

Table 8. Friedman test: ranking of methods.

Ranking	Method	Friedman's Score
1	HFs+DL + (ReliefF+bGWO)	2.40
1	HFs+DL + (ReliefF+GA)	2.40
2	HFs + (ReliefF+PSO)	3.60
2	HFs + (ReliefF+bGWO)	3.60
3	DL + (ReliefF+bGWO)	4.00
4	ResNet-50	6.20
5	EfficientNet-B2	6.80
6	DenseNet-121	7.20
7	VGG-19	8.60
8	Inception v3	9.40

Finally, we believe that the heterogeneous ensemble of classifiers was another relevant factor in achieving the results listed previously. We were able to define a combination of algorithms that supported the pattern recognition process of different types of H&E images with a more robust and reliable system capable of covering the weaknesses that may exist in a single classifier. In addition, we believe that the bias and variance have been reduced, minimizing the overfitting. More comparisons or algorithms could be considered to indicate the possible limits of each solution or even whether the main combinations are maintained from more descriptors or selection methods. However, the set of techniques with their associations and experiments described here provided an important overview of the potential and discriminative capacity regarding H&E-stained histology images.

Observations Based on Related Work

Considering the best results (Table 6), an overview is presented in relation to the literature involving each H&E-stained histological dataset. It is essential to note that even though the images share the same type, the datasets are not similar in the number of examples, classes, metrics used, validations, or distinct samples. Therefore, the purpose of this validation is not to compare performance rates directly. This situation requires equal conditions between the models: for example, by reproducing each strategy in the H&E-stained histological datasets explored here. This task is difficult or even infeasible. Hence, it is expected to observe whether the obtained solutions provide results among those available in the specialized literature. This illustrative overview is displayed in Tables 9–12 for the CR, OED, LA and LG, and NHL datasets, respectively.

From the values collected for this type of observation, it is verified that the solutions obtained here could provide highly competitive accuracy values, especially in three (CR, OED, and LG) of the five H&E datasets. In relation to the LA dataset (Table 11), the proposed approach provided an accuracy very close to the best solutions available in the literature: 99.24% (our solution) against 99.62% [29], which is a difference of only 0.38%. Taking into account the NHL dataset, even though the obtained solution indicated an accuracy among those achieved in the proposals of [62,109], a better-defined difference is verified in relation to the results of [20,29,56], showing that the strategies explored here deserve more attention for this type of image. Nevertheless, this overview has shown the discriminative capacity of the solutions obtained for different types of histological images, considering strategies not yet explored in the literature. Moreover, the ensemble learning approach provided relevant solutions with important information about the best combinations of descriptors and selection methods, using a reduced set of descriptors and revealing their main occurrences to recognize possible patterns in four types of histology tissues.

Table 9. Accuracy values provided by different strategies for CR image classification.

Author	Method	Accuracy
Proposed Method	DL + (ReliefF + bGWO)	100 %
[29]	ResNet50 with fine-tuning, multidimensional and multiscale fractal features	99.39%
[15]	ResNet50 (activation_48_relu layer), ReliefF and 35 deep-learned features	98.00%
[56]	8 CNN models, handcrafted descriptors	97.60%
[20]	9 CNN models, handcrafted descriptors	97.50%
[62]	Le-Net, multidimensional and multiscale fractal features, Haralick and LBPs	91.06%
[14]	ResNet deep-tuning (DL)	86.67%

Table 10. Accuracy rates achieved in different methods for OED image classification.

Author	Method	Accuracy
Proposed Method	HFs + (ReliefF + PSO)	100%
[17]	OralNet: Fused Optimal Deep Features	99.50%
[110]	Neural architecture search and handcrafted descriptors (morphological and non-morphological)	95.20%
[111]	Handcrafted descriptors (SIFT, SURF, ORB)	92.80%
[100]	Handcrafted descriptors (morphological and non-morphological)	92.40%
[16]	Densenet121	91.91%

Table 11. Accuracy rates indicated in different approaches for distinguishing LA and LG images.

Author	Method	Accuracy (LA)	Accuracy (LG)
Proposed Method	LA: HFs + (ReliefF + bGWO); LG: HFs+DL + (ReliefF + bGWO)	99.24%	100%
[29]	ResNet50 with fine-tuning, multidimensional and multiscale fractal features	99.62%	99.62%
[15]	ResNet50 (activation_48_relu layer), ReliefF and 5 deep-learned features	–	99.32%
[57]	Inception-V3, Fractal Dimension and Lacunarity (DL+HFs)	–	99.25%
[109]	CNN for texture	99.10%	98.20%
[112]	GIST handcrafted descriptor	88.40%	93.70%

Table 12. Accuracy values defined by different approaches for NHL image classification.

Author	Method	Accuracy
[20]	9 CNN models, handcrafted descriptors	97.33%
[56]	8 CNN models, handcrafted descriptors	97.33%
[29]	ResNet50 with fine-tuning, multidimensional and multiscale fractal features	95.55%
Proposed Method	HFs+DL + (ReliefF + GA)	92.25%
[62]	Le-Net, multidimensional and multiscale fractal features Haralick, LBPs	82.01%
[109]	CNN for texture	65.10%

4. Conclusions

In this work, an ensemble learning method was elaborated through multiple descriptors (handcrafted and deep-learned features), a two-stage feature selection, and a classification process with five algorithms (heterogeneous ensemble). The approach was utilized to categorize H&E histological images that are representative of various datasets, such as colorectal cancer, liver tissue, oral dysplasia, and non-Hodgkin's lymphomas.

The best ensembles indicated average accuracy values ranging from 90.72% (NHL) to 100% (CR). Since the initial feature set was composed of 11,086 values (462 handcrafted descriptors and 10,624 deep-learned features), the best solutions used a maximum of 53 features, with the following scenarios being noteworthy: CR with only 21 descriptors via bGWO; OED with only 11 descriptors via PSO; LA with 40 descriptors via bGWO; LG with only 29 attributes, through the bGWO; NHL with 53 descriptors, via GA. A breakdown of the main descriptors was also presented. It was observed that deep-learned descriptors predominated in relation to handcrafted ones, especially in the solutions for the LG and NHL datasets, with occurrences of 65.17% and 97.16%, respectively. On the other hand, the best solution for the LA dataset involved only handcrafted attributes. Another interesting behavior regarding handcrafted attributes is that the improved version of the Higuchi method outperformed the occurrences of important fractal techniques, specifically DF_n and DF_p, indicating the potential of the descriptor in multiple H&E-stained histological

datasets. In addition, the handcrafted features with the highest occurrences were lacunarity (45.80%, LA dataset), percolation (46.36%, OED dataset), LBPs (39.60%, LG dataset) and Haralick (100%, NHL dataset). The indications of solutions, attributes, and occurrences represent important contributions of this study, since the composition of each model and the conditions involved are available to specialists interested in these issues.

When comparing the optimal outcomes with those achieved through CNN architectures applied directly to the H&E-stained histological datasets, it is noted that the proposed approach presented a superior performance in all conditions explored here. Moreover, regarding the performances available in the specialized literature for the same image contexts, the proposal provided the best solutions in three (CR, OED, and LG) of the five datasets, exploring from 11 (OED) to 29 (LG) features. Therefore, these results confirm the proposal as a robust baseline approach capable of providing models without overfitting, offering valuable insights for the assessment and enhancement of CAD systems tailored explicitly for H&E samples, particularly those representing CR, OED, and LG.

Finally, some issues concerning the proposed approach deserve attention. For instance, the effectiveness of parameter tuning, algorithm inclusion, and attribute selection methods may heavily depend on the dataset explored. The solutions may not generalize well to other types of histological images. Also, the success of applying metaheuristics and other algorithms relies on their suitability for the given problem. Biases might arise if specific algorithms are more effective due to the nature of the data, potentially favoring certain types of classifiers. Finally, the use of cutoff points for attribute selection via the ReliefF algorithm introduces a subjective element. The chosen cutoff points could impact the definition of best attributes, leading to potential biases based on the selected thresholds.

In future work, we intend to investigate the following: the limits and impacts on the best ensembles after applying parameter tuning methods for metaheuristics, including other algorithms; a scheme that aims to understand why the features were selected, in addition to which of them are most important for the classification process; influences of cutoff points to define the best attributes via the ReliefF algorithm (first stage of selection); the discriminative power of handcrafted attributes and corresponding ensembles based on quantifications of explainable artificial intelligence representations, specifically gradient-weighted class activation mapping and locally interpretable model-agnostic explanations; the discriminative capacity of these combinations and conditions in other H&E-stained histological images; comparisons of the main results with other existing methods or algorithms commonly used in the analysis of histological images.

Author Contributions: Conceptualization, methodology, validation, formal analysis, investigation, writing—original draft preparation, writing—review and editing, J.J.T., L.H.d.C.L., G.F.R., T.A.A.T., P.R.d.F., A.M.L., S.V.C., A.B.S., M.Z.d.N. and L.A.N. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

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

Abbreviations

The following abbreviations are used in this manuscript:

H&E	hematoxylin and eosin
CAD	computer-aided diagnosis
HFs	handcrafted features
LBPs	local binary patterns
DL	deep-learned features
CNN	convolutional neural network
EL	ensemble learning
FC2	second fully connected layer
FC	last fully connected layer
GA	genetic algorithm
PSO	particle swarm optimization
bGWO	binary gray wolf optimization
SVM	support vector machine
KNN	K-nearest neighbors
AUC	area under the ROC curve
CR	colorectal cancer dataset
OED	oral epithelial dysplasia dataset
NHL	non-Hodgkin's lymphoma dataset
LG	liver gender dataset
LA	liver aging dataset
CLL	chronic lymphocytic leukemia
FL	follicular lymphoma
MCL	mantle cell lymphoma

References

1. Junqueira, L.C.; Carneiro, J. *Histologia Básica: Texto & Atlas*, 12th ed.; Guanabara Koogan: São Paulo, Brazil, 2013.
2. Nayak, S.R.; Mishra, J. An improved method to estimate the fractal dimension of colour images. *Perspect. Sci.* **2016**, *8*, 412–416. [CrossRef]
3. Frick, C.; Rumgay, H.; Vignat, J.; Ginsburg, O.; Nolte, E.; Bray, F.; Soerjomataram, I. Quantitative estimates of preventable and treatable deaths from 36 cancers worldwide: A population-based study. *Lancet Glob. Health* **2023**, *11*, e1700-12. [CrossRef] [PubMed]
4. Siegel, R.L.; Miller, K.D.; Fuchs, H.E.; Jemal, A. Cancer Statistics, 2021. *CA Cancer J. Clin.* **2021**, *71*, 7–33. [CrossRef]
5. Iftikhar, M.A.; Hassan, M.; Alquhayz, H. A colon cancer grade prediction model using texture and statistical features, SMOTE and mRMR. In Proceedings of the 2016 19th International Multi-Topic Conference (INMIC), Islamabad, Pakistan, 5–6 December 2016; pp. 1–7.
6. Akbar, B.; Gopi, V.P.; Babu, V.S. Colon cancer detection based on structural and statistical pattern recognition. In Proceedings of the 2015 2nd International Conference on Electronics and Communication Systems (ICECS), Coimbatore, India, 26–27 February 2015; pp. 1735–1739.
7. Altunbay, D.; Cigir, C.; Sokmensuer, C.; Gunduz-Demir, C. Color Graphs for Automated Cancer Diagnosis and Grading. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 665–674. [CrossRef] [PubMed]
8. ARAÚJO, T.; Aresta, G.; Castro, E.; Rouco, J.; Aguiar, P.; Eloy, C.; Polónia, A.; Campilho, A. Classification of breast cancer histology images using Convolutional Neural Networks. *PLoS ONE* **2017**, *12*, e0177544. [CrossRef]
9. Chan, H.P.; Hadjiiski, L.M.; Samala, R.K. Computer-aided diagnosis in the era of deep learning. *Med. Phys.* **2020**, *47*, e218–e227. [CrossRef]
10. Tang, J.; Rangayyan, R.M.; Xu, J.; Naqa, I.E.; Yang, Y. Computer-Aided Detection and Diagnosis of Breast Cancer With Mammography: Recent Advances. *IEEE Trans. Inf. Technol. Biomed.* **2009**, *13*, 236–251. [CrossRef]
11. JØrgensen, A.; Rasmussen, A.M.; Andersen, N.K.M.; Andersen, S.K.; Emborg, J.; RØge, R.; Østergaard, L. Using Cell Nuclei Features to Detect Colon Cancer Tissue in Hematoxylin and Eosin Stained Slides. *Cytom. Part A* **2017**, *91*, 785–793. [CrossRef]
12. Klonowski, W.; Pierzchalski, M.; Stepien, P.; Stepien, R.; Ahammer, H. Application of Higuchi's fractal dimension in analysis of images of Anal Intraepithelial Neoplasia. *Chaos Solitons Fractals* **2013**, *48*, 54–60. [CrossRef]
13. Ribeiro, M.G.; Neves, L.A.; Nascimento, M.Z.d.; Roberto, G.F.; Martins, A.M.; Tosta, T.A.A. Classification of colorectal cancer based on the association of multidimensional and multiresolution features. *Expert Syst. Appl.* **2019**, *120*, 262–278. [CrossRef]
14. Zhang, R.; Zhu, J.; Yang, S.; Hosseini, M.S.; Genovese, A.; Chen, L.; Rowsell, C.; Damaskinos, S.; Varma, S.; Plataniotis, K.N. HistoKT: Cross Knowledge Transfer in Computational Pathology. In Proceedings of the ICASSP 2022—2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, 23–27 May 2022; pp. 1276–1280.

15. de Oliveira, C.I.; do Nascimento, M.Z.; Roberto, G.F.; Tosta, T.A.; Martins, A.S.; Neves, L.A. Hybrid models for classifying histological images: An association of deep features by transfer learning with ensemble classifier. *Multimed. Tools Appl.* **2023**, 1–24. [CrossRef]

16. Maia, B.M.S.; de Assis, M.C.F.R.; de Lima, L.M.; Rocha, M.B.; Calente, H.G.; Correa, M.L.A.; Camisasca, D.R.; Krohling, R.A. Transformers, convolutional neural networks, and few-shot learning for classification of histopathological images of oral cancer. *Expert Syst. Appl.* **2023**, *241*, 122418. [CrossRef]

17. Mohan, R.; Rama, A.; Raja, R.K.; Shaik, M.R.; Khan, M.; Shaik, B.; Rajinikanth, V. OralNet: Fused Optimal Deep Features Framework for Oral Squamous Cell Carcinoma Detection. *Biomolecules* **2023**, *13*, 1090. [CrossRef]

18. Hassan, A.H.; Wahed, M.E.; Atiea, M.A.; Metwally, M.S. A hybrid approach for classification Breast Cancer histopathology Images. *Front. Sci. Res. Technol.* **2022**, *3*, 1–10. [CrossRef]

19. Nanni, L.; Ghidoni, S.; Brahnam, S. Handcrafted vs. non-handcrafted features for computer vision classification. *Pattern Recognit.* **2017**, *71*, 158–172. [CrossRef]

20. Nanni, L.; Ghidoni, S.; Brahnam, S.; Liu, S.; Zhang, L. Ensemble of Handcrafted and Deep Learned Features for Cervical Cell Classification. In *Deep Learners and Deep Learner Descriptors for Medical Applications*; Springer International Publishing: Cham, Switzerland, 2020; pp. 117–135. [CrossRef]

21. Sethy, P.K.; Behera, S.K. Automatic classification with concatenation of deep and handcrafted features of histological images for breast carcinoma diagnosis. *Multimed. Tools Appl.* **2022**, *81*, 9631–9643. [CrossRef]

22. Hu, W.; Li, X.; Li, C.; Li, R.; Jiang, T.; Sun, H.; Huang, X.; Grzegorzek, M.; Li, X. A state-of-the-art survey of artificial neural networks for whole-slide image analysis: From popular convolutional neural networks to potential visual transformers. *Comput. Biol. Med.* **2023**, *161*, 107034. [CrossRef] [PubMed]

23. Baish, J.W.; Jain, R.K. Fractals and Cancer. *Cancer Res.* **2000**, *60*, 3683–3688. [PubMed]

24. Ivanovici, M.; Richard, N. Fractal dimension of color fractal images, Image Processing. *IEEE Trans. Image Process.* **2011**, *20*, 227–235. [CrossRef] [PubMed]

25. Li, L.; Chang, L.; Ke, S.; Huang, D. Multifractal analysis and lacunarity analysis: A promising method for the automated assessment of muskmelon (*Cucumis melo* L.) epidermis netting. *Comput. Electron. Agric.* **2012**, *88*, 72–84. [CrossRef]

26. Lopes, R.; Betrouni, N. Fractal and multifractal analysis: A review. *Med. Image Anal.* **2009**, *13*, 634–649. [CrossRef]

27. Varley, T.F.; Craig, M.; Adapa, R.; Finoia, P.; Williams, G.; Allanson, J.; Pickard, J.; Menon, D.K.; Stamatakis, E.A. Fractal dimension of cortical functional connectivity networks & severity of disorders of consciousness. *PLoS ONE* **2020**, *15*, e0223812.

28. Tenguam, J.J.; Rozendo, G.B.; Roberto, G.F.; Nascimento, M.Z.; Martins, A.S.; Neves, L.A. Multidimensional and multiscale Higuchi dimension for the analysis of colorectal histological images. In Proceedings of the 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Seoul, Republic of Korea, 16–19 December 2020; pp. 2833–2839. [CrossRef]

29. Roberto, G.F.; Lumini, A.; Neves, L.A.; Nascimento, M.Z. Fractal Neural Network: A new ensemble of fractal geometry and convolutional neural networks for the classification of histology images. *Expert Syst. Appl.* **2021**, *166*, 114103. [CrossRef]

30. Haralick, R.M.; Shanmugam, K.; Dinstein, I. Textural Features for Image Classification. *IEEE Trans. Syst. Man Cybern.* **1973**, *SMC-3*, 610–621. [CrossRef]

31. Ojala, T.; PietikÄinen, M.; Harwood, D. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognit.* **1996**, *29*, 51–59. [CrossRef]

32. Almaraz-Damian, J.A.; Ponomaryov, V.; Sadovnychiy, S.; Castillejos-Fernandez, H. Melanoma and Nevus Skin Lesion Classification Using Handcraft and Deep Learning Feature Fusion via Mutual Information Measures. *Entropy* **2020**, *22*, 484. [CrossRef] [PubMed]

33. Nketiah, G.; Elschot, M.; Kim, E.; Teruel, J.R.; Scheenen, T.W.; Bathen, T.F.; Selnæs, K.M. T2-weighted MRI-derived textural features reflect prostate cancer aggressiveness: Preliminary results. *Eur. Radiol.* **2017**, *27*, 3050–3059. [CrossRef]

34. Wibmer, A.; Hricak, H.; Gondo, T.; Matsumoto, K.; Veeraraghavan, H.; Fehr, D.; Zheng, J.; Goldman, D.; Moskowitz, C.; Fine, S.W.; et al. Haralick texture analysis of prostate MRI: Utility for differentiating non-cancerous prostate from prostate cancer and differentiating prostate cancers with different Gleason scores. *Eur. Radiol.* **2015**, *25*, 2840–2850. [CrossRef]

35. Wang, C.; Yu, C. Automated morphological classification of lung cancer subtypes using H&E tissue images. *Mach. Vis. Appl.* **2013**, *24*, 1383–1391.

36. FONDÓN, I.; Sarmiento, A.; García, A.I.; Silvestre, M.; Eloy, C.; Polónia, A.; Aguiar, P. Automatic classification of tissue malignancy for breast carcinoma diagnosis. *Comput. Biol. Med.* **2018**, *96*, 41–51. [CrossRef]

37. Neves, L.A.; do Nascimento, M.; Oliveira, D.L.; Martins, A.S.; Godoy, M.F.; Arruda, P.F.; Neto, D.d.S.; Machado, J.M. Multi-scale lacunarity as an alternative to quantify and diagnose the behavior of prostate cancer. *Expert Syst. Appl.* **2014**, *41*, 5017–5029. [CrossRef]

38. Roberto, G.F.; Neves, L.A.; Nascimento, M.Z.; Tosta, T.A.A.; Longo, L.C.; Martins, A.S.; Faria, P.R. Features based on the percolation theory for quantification of non-Hodgkin lymphomas. *Comput. Biol. Med.* **2017**, *91*, 135–147. [CrossRef]

39. Klonowski, W.; Stepien, P.; Stepien, R.; Sedivy, R.; Ahammer, H.; Spasic, S. Analysis of Anal Intraepithelial Neoplasia Images using 1D and 2D Higuchi's fractal dimension methods. *Fractals* **2018**, *26*, 1850021. [CrossRef]

40. Roberto, G.F.; Nascimento, M.Z.; Martins, A.S.; Tosta, T.A.A.; Faria, P.R.; Neves, L.A. Classification of breast and colorectal tumors based on percolation of color normalized images. *Comput. Graph.* **2019**, *84*, 134–143. [CrossRef]

41. Ivanovici, M.; Richard, N.; Decean, H. Fractal Dimension and Lacunarity of Psoriatic Lesions—A Colour Approach. In Proceedings of the 2nd WSEAS International Conference on Biomedical Electronics and Biomedical Informatics, BEBI '09, Moscow, Russia, 20–29 August 2009; pp. 199–202.

42. Yu, Z.; Sohail, A.; Jamil, M.; Beg, O.; Tavares, J.M.R. Hybrid algorithm for the classification of fractal designs and images. *Fractals* **2023**, *31*, 1–11. [CrossRef]

43. Budnik, M.; Gutierrez-Gomez, E.; Safadi, B.; Pellerin, D.; Quénot, G. Learned features versus engineered features for multimedia indexing. *Multimed. Tools Appl.* **2017**, *76*, 11941–11958. [CrossRef]

44. Elmannai, H.; Hamdi, M.; AlGarni, A. Deep learning models combining for breast cancer histopathology image classification. *Int. J. Comput. Intell. Syst.* **2021**, *14*, 1003–1013. [CrossRef]

45. Greenspan, H.; Ginneken, B.; Summers, R.M. Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique. *IEEE Trans. Med Imaging* **2016**, *35*, 1153–1159. [CrossRef]

46. Jasti, V.D.P.; Zamani, A.S.; Arumugam, K.; Naved, M.; Pallathadka, H.; Sammy, F.; Raghuvanshi, A.; Kaliyaperumal, K. Computational Technique Based on Machine Learning and Image Processing for Medical Image Analysis of Breast Cancer Diagnosis. *Secur. Commun. Netw.* **2022**, *2022*, 1918379. [CrossRef]

47. Kassani, S.H.; Kassani, P.H.; Wesolowski, M.J.; Schneider, K.A.; Deters, R. Classification of histopathological biopsy images using ensemble of deep learning networks. In Proceedings of the CASCON '19: Proceedings of the 29th Annual International Conference on Computer Science and Software Engineering, Toronto, ON, Canada, 4–6 November 2019; pp. 92–99.

48. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.W.M.; Ginneken, B. Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [CrossRef] [PubMed]

49. Al-masni, M.A.; Al-antari, M.A.; Park, J.M.; Gi, G.; Kim, T.Y.; Rivera, P.; Valarezo, E.; Han, S.M.; Kim, T.S. Detection and classification of the breast abnormalities in digital mammograms via regional Convolutional Neural Network. In Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju Island, Republic of Korea, 11–15 July 2017; pp. 1230–1233.

50. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.Huang, Z. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]

51. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems 25*; Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2012; pp. 1097–1105.

52. 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]

53. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.

54. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

55. Rajinikanth, V.; Joseph Raj, A.; Thanaraj, K.; Naik, G. A Customized VGG19 Network with Concatenation of Deep and Handcrafted Features for Brain Tumor Detection. *Appl. Sci.* **2020**, *10*, 3429. [CrossRef]

56. Nanni, L.; Ghidoni, S.; Brahnam, S. Ensemble of convolutional neural networks for bioimage classification. *Appl. Comput. Inform.* **2018**, *17*, 19–35. [CrossRef]

57. Longo, L.H.D.C.; Martins, A.S.; Do Nascimento, M.Z.; Dos Santos, L.F.S.; Roberto, G.F.; Neves, L.A. Ensembles of fractal descriptors with multiple deep learned features for classification of histological images. In Proceedings of the 2022 29th International Conference on Systems, Signals and Image Processing (IWSSIP), Sofia, Bulgaria, 1–3 June 2022; pp. 1–4.

58. Zerouaoui, H.; Idri, A.; El Alaoui, O. A new approach for histological classification of breast cancer using deep hybrid heterogenous ensemble. *Data Technol. Appl.* **2023**, *57*, 245–278. [CrossRef]

59. Hagerty, J.R.; Stanley, R.J.; Almubarak, H.A.; Lama, N.; Kasmi, R.; Guo, P.; Drugge, R.J.; Rabinovitz, H.S.; Oliviero, M.; Stoecker, W.V. Deep Learning and Handcrafted Method Fusion: Higher Diagnostic Accuracy for Melanoma Dermoscopy Images. *IEEE J. Biomed. Health Inform.* **2019**, *23*, 1385–1391. [CrossRef]

60. Ponti, M.P., Jr. Combining Classifiers: From the Creation of Ensembles to the Decision Fusion. In Proceedings of the 2011 24th SIBGRAPI Conference on Graphics, Patterns, and Images Tutorials, Maceió, Brazil, 28–31 August 2011; pp. 1–10. [CrossRef]

61. Bhowal, P.; Sen, S.; Sarkar, R. A two-tier feature selection method using Coalition game and Nystrom sampling for screening COVID-19 from chest X-ray images. *J. Ambient. Intell. Humaniz. Comput.* **2021**, *14*, 3659–3674 [CrossRef]

62. Candelero, D.; Roberto, G.F.; do Nascimento, M.Z.; Rozendo, G.B.; Neves, L.A. Selection of CNN, Haralick and Fractal Features Based on Evolutionary Algorithms for Classification of Histological Images. In Proceedings of the 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Seoul, Republic of Korea, 16–19 December 2020; pp. 2709–2716.

63. Li, Y.; Shen, Y.; Fan, X.; Huang, X.; Yu, H.; Zhao, G.; Ma, W. A novel EEG-based major depressive disorder detection framework with two-stage feature selection. *BMC Med. Inform. Decis. Mak.* **2022**, *22*, 209. [CrossRef]

64. Zhang, X.; Zhang, Q.; Chen, M.; Sun, Y.; Qin, X.; Li, H. A two-stage feature selection and intelligent fault diagnosis method for rotating machinery using hybrid filter and wrapper method. *Neurocomputing* **2018**, *275*, 2426–2439. [CrossRef]

65. Li, X.; Wu, J.; Jiang, H.; Chen, E.Z.; Dong, X.; Rong, R. Skin Lesion Classification Via Combining Deep Learning Features and Clinical Criteria Representations. *bioRxiv* **2018**. [CrossRef]

66. Kumar, N.; Sharma, M.; Singh, V.P.; Madan, C.; Mehandia, S. An empirical study of handcrafted and dense feature extraction techniques for lung and colon cancer classification from histopathological images. *Biomed. Signal Process. Control* **2022**, *75*, 103596. [CrossRef]

67. Zhao, X.; Li, D.; Yang, B.; Chen, H.; Yang, X.; Yu, C.; Liu, S. A two-stage feature selection method with its application. *Comput. Electr. Eng.* **2015**, *47*, 114–125. [CrossRef]

68. Nikolaidis, N.S.; Nikolaidis, I.N.; Tsouros, C.C. A Variation of the Box-Counting Algorithm Applied to Colour Images. *arXiv* **2011**, arXiv:1107.2336.

69. Strelniker, Y.M.; Havlin, S.; Bunde, A. Fractals and Percolation. In *Encyclopedia of Complexity and Systems Science*; Springer: New York, NY, USA, 2009; pp. 3847–3858.

70. Yamaguchi, T.; Hashimoto, S. Fast crack detection method for large-size concrete surface images using percolation-based image processing. *Mach. Vis. Appl.* **2010**, *21*, 797–809. [CrossRef]

71. Higuchi, T. Approach to an irregular time series on the basis of the fractal theory. *Phys. D Nonlinear Phenom.* **1988**, *31*, 227–283. [CrossRef]

72. Gomes, R.L.; Vanderlei, L.C.M.; Garner, D.M.; Vanderlei, F.M.; Valenti, V.E. Higuchi Fractal Analysis of Heart Rate Variability is Sensitive during Recovery from Exercise in Physically Active Men. *MedicalExpress* **2017**, *4*, 1–8.

73. Gomolka, R.S.; Kampusch, S.; Kaniusas, E.; Thürk, F. Széles, J.C.; Klonowski, W. Higuchi Fractal Dimension of Heart Rate Variability During Percutaneous Auricular Vagus Nerve Stimulation in Healthy and Diabetic Subjects. *Front. Physiol.* **2018**, *9*, 1162. [CrossRef]

74. Ivanovici, M.; Richard, N. The lacunarity of colour fractal images. In Proceedings of the 16th IEEE International Conference on Image Processing (ICIP), Cairo, Egypt, 7–10 November 2009; pp. 453–456.

75. Căliman, A. ; Ivanovici, M. Psoriasis image analysis using color lacunarity. In Proceedings of the 2012 13th International Conference on Optimization of Electrical and Electronic Equipment (OPTIM), Brasov, Romania, 24–26 May 2012; pp. 1401–1406.

76. Dinčić, M.; Todorović, J.; Nešović Ostojić, J.; Kovačević, S.; Dunđerović, D.; Lopičić, S.; Spasić, S.; Radojević-Škodrić, S.; Stanisavljević, D.; Ilić, A.Ž. The fractal and GLCM textural parameters of chromatin may be potential biomarkers of papillary thyroid carcinoma in Hashimoto's thyroiditis specimens. *Microsc. Microanal.* **2020**, *26*, 717–730. [CrossRef]

77. Korkmaz, S.A.; Binol, H. Classification of molecular structure images by using ANN, RF, LBP, HOG, and size reduction methods for early stomach cancer detection. *J. Mol. Struct.* **2018**, *1156*, 255–263. [CrossRef]

78. Yu, C.; Chen, H.; Li, Y.; Peng, Y.; Li, J.; Yang, F. Breast cancer classification in pathological images based on hybrid features. *Multimed. Tools Appl.* **2019**, *78*, 21325–21345. [CrossRef]

79. Mazo, C.; Alegre, E.; Trujillo, M. Classification of cardiovascular tissues using LBP based descriptors and a cascade SVM. *Comput. Methods Programs Biomed.* **2017**, *147*, 1–10. [CrossRef]

80. MathWorks. Extract Local Binary Pattern (LBP) Features. 2022. Available online: https://www.mathworks.com/help/vision/ref/extractlbpfeatures.html#buumhti-1-CellSize (accessed on 18 May 2022).

81. Torrey, L.; Shavlik, J. Transfer Learning. In *Handbook of Research on Machine Learning Applications*; Soria, E., Martin, J., Magdalena, R., Martinez, M., Serrano, A., Eds.; IGI Global: Hershey, PA, USA, 2009.

82. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [CrossRef]

83. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.

84. Huang, G.; Liu, Z.; Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269. [CrossRef]

85. Tan, M.; Le, Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; Volume 97, pp. 6105–6114.

86. Rakhlin, A.; Shvets, A.; Iglovikov, V.; Kalinin, A.A. Deep Convolutional Neural Networks for Breast Cancer Histology Image Analysis. *Image Anal. Recognition. ICIAR 2018* **2018**, *10882*, 737–744.

87. Kwasigroch, A.; Mikołajczyk, A.; Grochowski, M. Deep neural networks approach to skin lesions classification—A comparative analysis. In Proceedings of the 2017 22nd International Conference on Methods and Models in Automation and Robotics (MMAR), Miedzyzdroje, Poland, 28–31 August 2017; pp. 1069–1074.

88. dos Santos, F.P.; Ponti, M.S. Alignment of Local and Global Features from Multiple Layers of Convolutional Neural Network for Image Classification. In Proceedings of the 2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), Rio Grande, Brazil, 28–30 October 2019; pp. 241–248.

89. Ben Hamida, A.; Devanne, M.; Weber, J.; Truntzer, C.; Derangère, V.; Ghiringhelli, F.; Forestier, G.; Wemmert, C. Deep learning for colon cancer histopathological images analysis. *Comput. Biol. Med.* **2021**, *136*, 104730. [CrossRef] [PubMed]

90. Kalra, S.; Tizhoosh, H.R.; Choi, C.; Shah, S.; Diamandis, P.; Campbell, C.J.V.; Pantanowitz, L. Yottixel—An Image Search Engine for Large Archives of Histopathology Whole Slide Images. *Med Image Anal.* **2020**, *65*, 101757. [CrossRef] [PubMed]

91. Munien, C.; Viriri, S. Classification of Hematoxylin and Eosin-Stained Breast Cancer Histology Microscopy Images Using Transfer Learning with EfficientNets. *Comput. Intell. Neurosci.* **2021**, *2021*, 5580914. [CrossRef]

92. Wolfram. Wolfram Neural Net Repository. 2022. Available online: https://resources.wolframcloud.com/NeuralNetRepository/ (accessed on 18 August 2022).

93. Hsu, H.; Hsieh, C.W.; Lu, M. Hybrid feature selection by combining filters and wrappers. *Expert Syst. Appl.* **2011**, *38*, 8144–8150. [CrossRef]

94. Mengdi, L.; Liancheng, X.; Jing, Y.; Jie, H. A Feature Gene Selection Method Based on ReliefF and PSO. In Proceedings of the 2018 10th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), Changsha, China, 10–11 February 2018; pp. 298–301.

95. Urbanowicz, R.J.; Meeker, M.; La Cava, W.; Olson, R.S.; Moore, J.H. Relief-based feature selection: Introduction and review. *J. Biomed. Inform.* **2018**, *85*, 189–203. [CrossRef] [PubMed]

96. Emarya, E.; Zawbaab, H.M.; Hassaniena, A.E. Binary grey wolf optimization approaches for feature selection. *Neurocomputing* **2016**, *172*, 371–381. [CrossRef]

97. Taino, D.T.; Ribeiro, M.G.; Roberto, G.F.; Zafalon, G.F.D.; Do Nascimento, M.Z.; Tosta, T.A.A.; Martins, A.S.; Neves, L.A. Analysis of cancer in histological images: Employing an approach based on genetic algorithm. *Pattern Anal. Appl.* **2021**, *24*, 483–496. [CrossRef]

98. Bradley, A.P. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern Recognit.* **1997**, *30*, 1145–1159. [CrossRef]

99. Sirinukunwattana, K.; Pluim, J.P.W.; Chen, H.; Qi, X. Gland segmentation in colon histology images: The glas challenge contest. *MedicalImageAnalysis* **2017**, *35*, 489–502. [CrossRef]

100. Silva, A.B.; Martins, A.S.; Tosta, T.A.A.; Neves, L.A.; Servato, J.P.S.; de Araújo, M.S.; De Faria, P.R.; Do Nascimento, M.Z. Computational analysis of histological images from hematoxylin and eosin-stained oral epithelial dysplasia tissue sections. *Expert Syst. Appl.* **2022**, *193*, 116456. [CrossRef]

101. Shamir, L.; Orlov, N.; Eckley, D.M.; Macura, T.J.; Goldberg, I.G. IICBU 2008: A proposed benchmark suite for biological image analysis. *Med. Biol. Eng. Comput.* **2008**, *46*, 943–947. [CrossRef] [PubMed]

102. Zahn, J.M.; Poosala, S.; Owen, A.B.; Ingram, D.K.; Lustig, A.; Carter, A.; Weeraratna, A.T.; Taub, D.D.; Gorospe, M.; Mazan-Mamczarz, K.; et al. AGEMAP: A Gene Expression Database for Aging in Mice. *PLoS Genet.* **2007**, *3*, e201. [CrossRef]

103. Mathworks. R2019a at a Glance. 2021. Available online: https://ch.mathworks.com/solutions/deep-learning/models.html (accessed on 20 July 2021).

104. Pytorch. 2021. Available online: https://pytorch.org (accessed on 20 July 2022).

105. Google. Colaboratory, 2021. Available online: https://research.google.com/colaboratory/faq.html (accessed on 20 July 2021).

106. Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I.H. The WEKA data mining software: An update. *ACM SIGKDD Explor. Newsl.* **2009**, *11*, 10–18. [CrossRef]

107. Pytorch. Models and Pre-trained Weights. 2022. Available online: https://pytorch.org/vision/stable/models.html (accessed on 20 July 2021).

108. Demšar, J. Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learn. Res.* **2006**, *7*, 1–30.

109. Andrearczyk, V.; Whelan, P.F. Deep learning for biomedical texture image analysis. In Proceedings of the Irish Machine Vision & Image Processing Conference. Irish Pattern Recognition & Classification Society (IPRCS), Maynooth, Ireland, 30 August–1 September 2017.

110. Azarmehr, N.; Shephard, A.; Mahmood, H.; Rajpoot, N.; Khurram, S.A. Automated oral epithelial dysplasia grading using neural networks and feature analysis. In Proceedings of the Medical Imaging with Deep Learning, Zurich, Switzerland, 6–8 July 2022.

111. Adel, D.; Mounir, J.; El-Shafey, M.; Eldin, Y.A.; Masry, N.E.; Abdelraouf, A.; Elhamid, I.S.A. Oral Epithelial Dysplasia Computer Aided Diagnostic Approach. In Proceedings of the 2018 13th International Conference on Computer Engineering and Systems (ICCES), Cairo, Egypt, 18–19 December 2018; Volume 2, pp. 313–318. [CrossRef]

112. Watanabe, K.; Kobayashi, T.; Wada, T. Semi-supervised feature transformation for tissue image classification. *PLoS ONE* **2016**, *11*, e0166413. [CrossRef]