Detection of Covid-19 in Chest X-Ray Images Using Percolation Features and Hermite Polynomial Classification

Published Version:
F. Roberto, G., C. Pereira, D., S. Martins, A., A. A. Tosta, T., Soares, C., Lumini, A., et al. (2024). Detection of Covid-19 in Chest X-Ray Images Using Percolation Features and Hermite Polynomial Classification. Cham : Springer [10.1007/978-3-031-49018-7_12].

Availability:
This version is available at: https://hdl.handle.net/11585/959152 since: 2024-02-19

Published:
DOI: http://doi.org/10.1007/978-3-031-49018-7_12

Detection of Covid-19 in Chest X-Ray Images Using Percolation Features and Hermite Polynomial Classification

Guilherme F. Roberto[1(✉)], Danilo C. Pereira[2], Alessandro S. Martins[2], Thaína A. A. Tosta[3], Carlos Soares[1], Alessandra Lumini[4], Guilherme B. Rozendo[4,5], Leandro A. Neves[5], and Marcelo Z. Nascimento[6]

[1] Faculty of Engineering, University of Porto (FEUP), Porto, Portugal
guilhermefr@fe.up.pt
[2] Federal Institute of Education, Science and Technology of Triângulo Mineiro (IFTM), Ituiutaba-MG, Brazil
[3] Science and Technology Institute, Federal University of São Paulo (UNIFESP), São José dos Campos-SP, Brazil
[4] Department of Computer Science and Engineering (DISI), University of Bologna, Cesena, Italy
[5] Department of Computer Science and Statistics (DCCE), São Paulo State University (UNESP), São José do Rio Preto-SP, Brazil
[6] Faculty of Computer Science (FACOM), Federal University of Uberlândia (UFU), Uberlândia-MG, Brazil

Abstract. Covid-19 is a serious disease caused by the Sars-CoV-2 virus that has been first reported in China at late 2019 and has rapidly spread around the world. As the virus affects mostly the lungs, chest X-rays are one of the safest and most accessible ways of diagnosing the infection. In this paper, we propose the use of an approach for detecting Covid-19 in chest X-ray images through the extraction and classification of local and global percolation-based features. The method was applied in two datasets: one containing 2,002 segmented samples split into two classes (Covid-19 and Healthy); and another containing 1,125 non-segmented samples split into three classes (Covid-19, Healthy and Pneumonia). The 48 obtained percolation features were given as input to six different classifiers and then AUC and accuracy values were evaluated. We employed the 10-fold cross-validation method and evaluated the lesion sub-types with binary and multiclass classification using the Hermite Polynomial classifier, which had never been employed in this context. This classifier provided the best overall results when compared to other five machine learning algorithms. These results based in the association of percolation features and Hermite polynomial can contribute to the detection of the lesions by supporting specialists in clinical practices.

Keywords: Percolation · Chest X-ray images · Covid-19 · Handcrafted features · Computer vision

1 Introduction

The novel coronavirus (Sars-CoV-2) has been first reported in Wuhan, at the Chinese province of Hubei at the end of the year 2019 and has rapidly spread to other countries and become a worldwide pandemic, having the World Health Organization (WHO) declare a global emergency [25]. The virus infects lung alveolar epithelial cells, causing the disease named Covid-19, wherein its initial clinical sign is often pneumonia, although gastrointestinal symptoms and asymptomatic infections have also been reported [29].

According to the Coronavirus Research Center of the Johns Hopkins University, over 545 million cases have already been reported worldwide, which have so far resulted in more than 6.88 million deaths [6]. The conventional diagnosis is made through a technique called reverse transcription-polymerase chain reaction (RT-PCR) [4]. However, this procedure relies on the existence of material kits, which can be costly, and there is also a contamination risk due to the proximity between medical staff and the patient at the moment of the exam [17]. Medical imaging techniques such as chest X-rays and CT-scans have also performed a major role in this pandemic, as the virus usually infects the lungs, the images obtained from these exams can be evaluated by radiologists in order to verify whether the patient is infected with Sars-CoV-2 and its severity. Besides being safe and easily accessible, signs of Covid-19 could be observed in chest X-ray images even in asymptomatic cases [1,5].

Since the start of the pandemic, the work required to evaluate all incoming chest X-ray images has become time-consuming and prone to inaccuracies, as misdiagnosis can happen due to lack of experience of some radiologists in not yet having deep knowledge of how a Sars-CoV-2 infection evolves [17]. Therefore, Artificial Intelligence tools, mostly based on deep learning, have been developed recently in order to reduce the workload of radiologists and provide support to the diagnosis of Covid-19 [9,20,27,28]. Due to good performances obtained in the detection of pneumonia in chest X-ray images, convolutional neural networks (CNN) have promptly been applied to the detection of Covid-19 as soon as the first chest X-ray datasets were available [9,17,20,27].

One of the earliest frameworks designed for the detection of Covid-19 was COVIDX-Net [9], which included seven different CNN models and was able to obtain accuracy values up to 90%. However, this framework had limited training since at the time of its publication, only 25 Covid-19 X-ray image samples were available. As the number of Covid-19 samples in public datasets began to grow, more complex solutions were proposed. In [20] a 17-layer CNN named DarkCovidNet was proposed. This model was able to obtain an accuracy of 98.08% for binary classification (Covid-19 vs. No findings) and 87.02% for multiclass classification (Covid-19 vs. No findings vs. Pneumonia), which was later improved to 99.27% by [27] after the application of the Fuzzy Color approach on the preprocessing stage and the use of two different CNN models: MobileNetV2 and SqueezeNet. Besides, other complex CNN models have been used for this task. In [12], InceptionV3, Xception and ResNeXt models have been evaluated on a dataset containing 6,432 samples of chest X-rays from Covid-19, Pneumonia and Healthy cases. After splitting the dataset into training (85%) and validation (15%) subsets the models were evaluated separately. The model that provided the highest accuracy was the Xception, with an value of 97.97%. Recently, a modified version of the DarkCovidNet which contains two extra convolutional layers was able to obtain an accuracy of 94.18% for multiclass classification [22], although the number of samples in the dataset has significantly increased from when the first DarkCovidNet model was published.

However, it can be noted that the majority of approaches proposed for detecting Covid-19 is purely based on deep learning techniques, and although CNN models became popular due to the high accuracy rates they have provided on several types of image datasets, they are also known to have a significant higher computational cost than more traditional techniques, which can provide features known as handcrafted. Few researchers have chosen this approach so far, and these have shown that handcrafted features can perform as well as learned features in the detection of Covid-19, as they are often good indicators of shape, size and regularity of structures present in medical images [13]. For instance, in [28], the authors presented a simple local binary pattern (LBP) extractor to obtain features from a chest X-ray dataset composed by 87 Covid-19 samples and 147 healthy samples. The most relevant features were selected using the ReliefF feature selector algorithm and an average accuracy of 99.55% was obtained using an SVM classifier and 10-fold cross validation. In [11], preprocessing multiresolution approaches such as shearlet transform were applied to the chest X-ray images prior to feature extraction. Then, entropy and energy features were obtained from the multiresolution images and given as input for an extreme learning machine (ELM) classifier. The average accuracy obtained was of 99.29%. Other classical handcrafted features such as Haralick's have been used in computer vision problems regarding the detection of Covid-19, providing accuracy values up to 98.04% [7,16]. Apart from the computational cost, researchers often choose handcrafted features as these often perform better on small datasets, as indicated in [19], wherein the authors obtained robust features from 260 images using four different feature extraction techniques. Then, stacked auto-encoder (SAE) and principal component analysis (PCA) were applied to reduce the number of features to 20. These features were given as input to a SVM classifier which provided an accuracy of 94.23%.

Although these works have presented great results, there seems to be a lack of information on how other types of handcrafted features could perform. Therefore, we have decided to investigate the performance of percolation features for detecting Covid-19 in chest X-ray images. These type of features are fractal-based and have recently been applied for the classification of colored histology images [23]. However, percolation features have not yet been applied for the classification of X-ray images, which are greyscale. In fact, fractal descriptors for classifying Covid-19 images have only been used so far in combination with CNN [8]. Besides feature extraction through percolation properties, we also propose the use of a polynomial classifier. Although several methods contribute in solving biomedical classification problems, the choice of the appropriate technique for this task depends on the classifier properties [21]. Polynomial classifiers have been showing promising results, especially when dealing with non-linearly separable data. This algorithm is a parameterized method that expands exponentially its polynomial basis according to the number of elements in the vector of characteristics and the degree of the function. One of the most recent polynomial classifiers available is the Hermite polynomial (HP), which consists on an expansion of the regular polynomial basis to a larger space in order to analyze complex contexts such as the one explored here [15].

We propose in this paper an approach for detecting Covid-19 in chest X-ray images based on local and global percolation features, using an adaptation of the method described in [23]. These features were then classified by the HP algorithm that allows the evaluation of data distribution of images from different classes. This is the first time wherein these features are used to describe properties of chest X-ray images. Another contribution of our work is that it defines the best association after applying well-known algorithms of artificial intelligence to evaluate the performance of the proposed approach.

The contributions of this paper can be summarised as follows: investigation of the behaviour of local and global percolation features for the feature extraction from chest X-ray images; association of an approach based on the HP algorithm and percolation features capable of evaluating classes in binary classifications and; a comparison of well-known algorithms of artificial intelligence with the proposed feature extraction approach to evaluate the classification performance.

This paper is organised as follows. In Sect. 2, we detail the methodologies used in the various stages of the proposed approach and the datasets used in this paper. We discuss the obtained results in Sect. 3 and we conclude the paper on Sect. 4.

2 Materials and Methods

2.1 Image Database

V7 Labs Dataset. The first analyzed dataset was made available by [14]. This dataset is composed of 2,002 chest X-ray images split into two classes (401 Covid-19 infections and 1,601 healthy lung images) with dimensions ranging from 148×122 to 4030×3133. Lung segmentation was performed by human-made annotations. Images containing graphic indicators such as arrows inside the lung area were disregarded. In Fig. 1, one example from each class and their respective segmented images are presented.

Ozturk *et al.* Dataset. The second analyzed dataset consists in a composition made by [20] from two other datasets: one containing 125 Covid-19 samples; and other composed of 500 pneumonia samples and 500 samples with no findings. Therefore, this datasets consists of a multiclass classification problem. There were no annotations available for lung segmentation in this dataset, so the images were used as they were originally available in the repository. Examples from each class of this dataset are shown in Fig. 2.

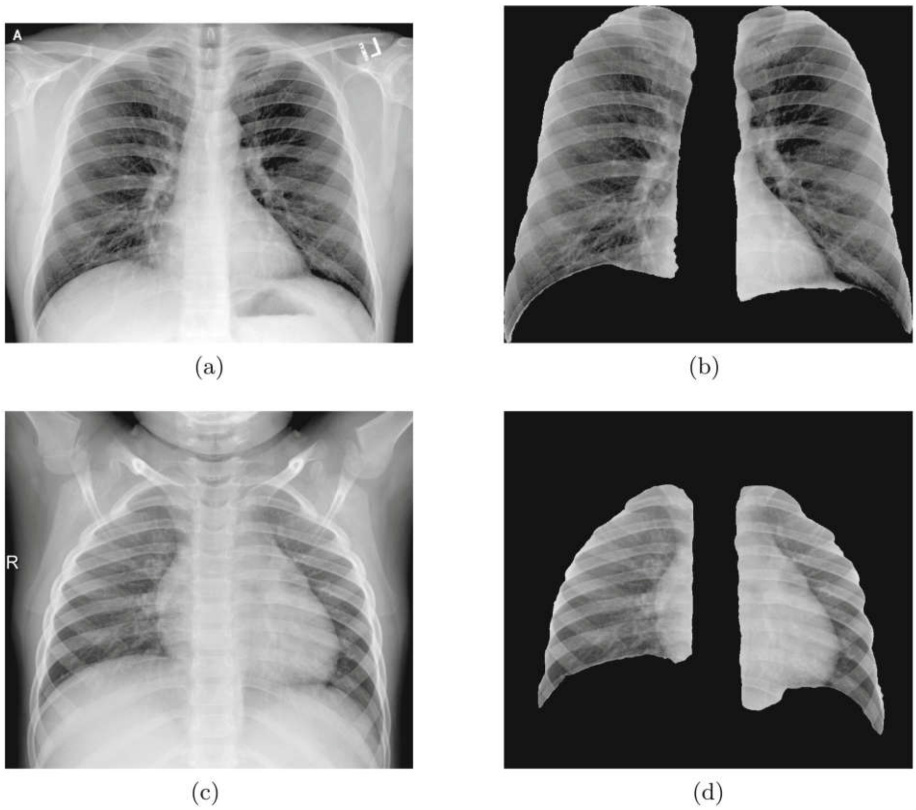


Fig. 1. Images from the v7 Labs dataset. (a) chest X-ray from an individual infected with Sars-CoV-2; (b) segmented version of the image in (a); (c) chest X-ray from an individual with a healthy lung; (d) segmented version of the image in (c).

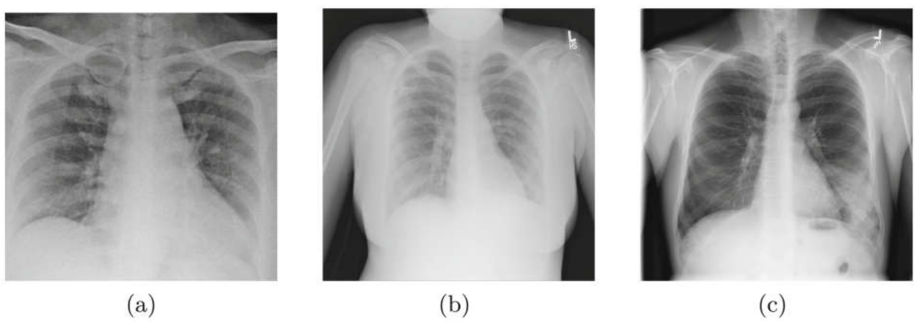


Fig. 2. Images from the Ozturk *et. al* dataset. (a) sample of a Covid-19 case; (b) sample of a healthy case; (c) sample of a pneumonia case.

2.2 Methodology

The approach proposed in this paper consists in applying the feature extraction algorithm available at [23] to obtain numeric features from the X-ray images from both datasets and then providing these features as input to a classifier. The overview of our approach is presented in Fig. 3.

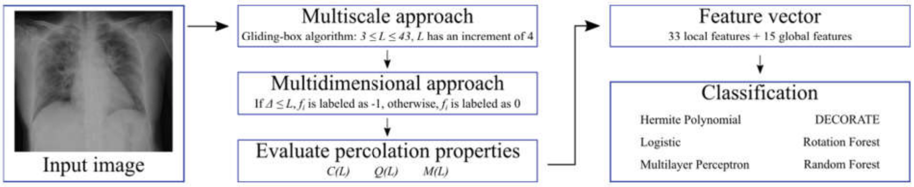


Fig. 3. Overview of the applied methodology.

Feature Extraction. The algorithm applied to obtain the percolation features consists in a multiscale and multidimensional method based on the analysis of the image under different scales valued L, wherein L ranges from 3 to 43 with an increment of 4, which are the parameters that provided the best results in [23]. This analysis is made through the application of the gliding-box algorithm, wherein for each iteration, a box sized $L \times L$ overlapping the image glides from the top left corner until it reaches the bottom right corner. The next steps of the algorithm are applied inside this box and the final features are calculated from an average of all the boxes of a given scale L.

Three percolation properties are verified: average number of clusters ($C(L)$); occurrence of percolation ($Q(L)$) and; average size of the largest cluster ($M(L)$). In the original paper, these properties were obtained after calculating the distance Δ between the RGB intensities of the center pixel f_c and any other given pixel f_i on each of the boxes sized $L \times L$. Since the chest X-ray images on both datasets are greyscale, we have adapted the distance calculation formula to the presented in Eq. 1:

$$\Delta = |f_i - f_c| \,. \tag{1}$$

This procedure outputs a set of 33 local features based on the three described percolation properties, and 15 global features based on the metrics defined by area (A), skewness (S), area ratio (A_R), maximum point (Max) and scale of the maximum point (σ). The global features were obtained using the approach proposed by [2]. A summary of the 48 obtained features can be observed on Table 1.

Table 1. Summary of the 48 extracted features.

Local Features (33)			Global Features (15)		
$C(3)$	$Q(3)$	$M(3)$	$A(C)$	$A(Q)$	$A(M)$
$C(7)$	$Q(7)$	$M(7)$	$S(C)$	$S(Q)$	$S(M)$
$C(11)$	$Q(11)$	$M(11)$	$A_R(C)$	$A_R(Q)$	$A_R(M)$
$\vdots$	$\vdots$	$\vdots$	$Max(C)$	$Max(Q)$	$Max(M)$
$C(43)$	$Q(43)$	$M(43)$	$\sigma(C)$	$\sigma(Q)$	$\sigma(M)$

Classification. One of the main contributions of this paper consists on the evaluation of percolation features obtained from chest X-ray images when given as input to a polynomial classifier.

The polynomial classifier is a supervised method capable of non-linearly expanding an input feature vector into a higher dimension. This strategy allows one to obtain linear approximations in this space that is capable of labeling the input data for a desired output [24]. However, regular polynomial classifiers have a limited basis, which hinders the evaluation of the features.

Hermite polynomials (HP) are able to generate a complete orthogonal basis of Hilbert space that satisfies the orthogonality and completeness conditions of the family of elements of that space [10]. The orthogonality condition, Eq. (2), implies that any inner product between a pair of orthogonal polynomials, $P_m(x)_{n=0}^{\infty}$ and $P_n(x)_{n=0}^{\infty}$, of different degrees is equal to zero when the base function used $w(x)$ and interval $[a, b]$ are the same [3].

$$< P_m(x), P_n(x) >= \int_a^b P_m(x)P_n(x)w(x)dx = 0, \forall m \neq n. \quad (2)$$

According to [26], HP is orthogonal in the interval $(-\infty, \infty)$ with respect to the base-function and enables gains at function approximation. Mathematically, HP can be defined as Eq. (3):

$$H_n(x) = (-1)^n e^{-x^2/2} \frac{d^n}{dx^n}[e^{-x^2/2}]. \quad (3)$$

A computationally efficient way to obtain the polynomial is to use the recurrence relation. In this case, only the first two terms are needed, so the others can be calculated iteratively. HP recurrence relation is expressed by Eq. (4) [30].

$$H_{n+1}(x) - xH_n(x) + nH_{n-1}(x) = 0, \quad n = 0, 1, 2, \ldots \quad (4)$$

Finally, the polynomial classifier can be defined according to Eq. (5):

$$g(\mathbf{x}) = \mathbf{a}^T H_n(\mathbf{x}), \quad (5)$$

where $\mathbf{a}$, is the vector of coefficients at polynomial basis function, $H_n(\mathbf{x})$ is the base-function of Hermite and n corresponds to the order or degree of the polynomial function.

The polynomial classifier requires two steps for operation: training and testing. In the first stage, the training data vector is expanded through polynomial expansion to increase the separation of the different classes in the newly generated vector space. In the second part, a mapping of the expanded polynomial vectors to an optimal output sequence to minimize an objective criterion is performed. The mapped parameters represent the weights of the polynomial classifier which are known as class models [18].

In this paper, the 48 obtained features were given as input to a classifier in order to predict the class which the image represents. The order of the base-polynomial was set empirically and the best results were achieved with the $4th$ order and four features for class separation. In addition, the cross-validation method with 10-folds was employed, where 90% of the dataset was used for training and 10% on model prediction and evaluation.

Performance Evaluation. We evaluated the performance of our proposed approach in three stages. Firstly, we evaluated the features by applying the Mann-Whitney U test [15]. Then, the HP classifier was evaluated by comparing the performance of global and local features separately and then on a single feature vector. The performance of the HP algorithm was also compared with other five different classifiers that are based on three of the main supervised machine learning approaches: function-based, ensemble learning and tree-based. The chosen algorithms were logistic (LGT), multilayer perceptron (MLP), DECORATE (DEC), rotation forest (RoF) and random forest (RaF).

All classifications were performed in the software Weka v3.6.13 using the default parameters for the classifiers and a 10-fold cross validation approach. The metrics used for evaluation were the accuracy (ACC) as it is the most common metric and has been used in every related work previously presented; and the area under the ROC curve (AUC), since it is a more suitable metric to evaluate performance with imbalanced data.

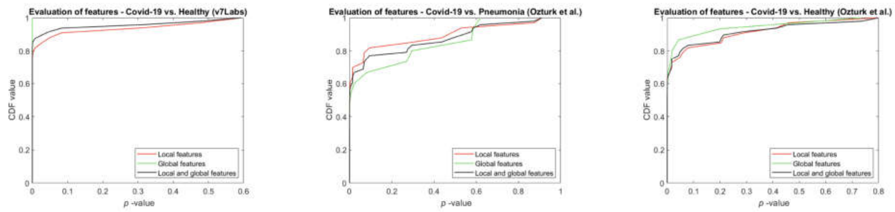
All tests were performed on an Intel i5-9300H at 2.40GHz and 8GB of RAM. The percolation features were obtained using Matlab R2019a.

3 Results and Discussion

3.1 Feature Evaluation

In order to evaluate the significance of the features and how discriminating they are on the different tested scenarios, we applied the Mann-Whitney U test [15]. The results are shown in Fig. 4, with the empirical cumulative distribution function (CDF) for the p-values. For the v7 Labs dataset, all of the 15 global features had p-values close to zero, as seen on Fig. 4a. As for the local features, approximately 85% of these had p-values smaller than 0.1.

For the Ozturk *et al.* dataset, the two classification scenarios were evaluated separately. In the first one, presented in Fig. 4b, the local features have provided slightly more significant results, with p-values smaller than 0.2 for 80% of the features. On the other hand, in the second scenario presented in Fig. 4c, the smallest p-values were provided by the global features, as approximately 80% of these resulted in p-values close to zero. For better understanding how these features behave when given as input to the HP classifier, we performed the evaluations described in the following subsection.



(a) (b) (c)

Fig. 4. The empirical cumulative distribution function of the p-values of the different feature sets: (a) v7Labs dataset; (b) Ozturk *et al.* dataset (Covid-19 vs. Pneumonia); (c) Ozturk *et al.* dataset (Covid-19 vs. Healthy).

3.2 Performance of the HP Classifier

We first evaluate the HP classifier using different orders for the polynomial function. This evaluation was made using the full set of global and local features in both Covid-19 vs. Healthy (CH), Covid-19 vs. Pneumonia (CP) and Covid-19 vs. Healthy vs. Pneumonia (CHP) scenarios. The results obtained for 1st (HPG1), 2nd (HPG2), 3rd (HPG3) and 4th (HPG4) orders are shown in Fig. 5.

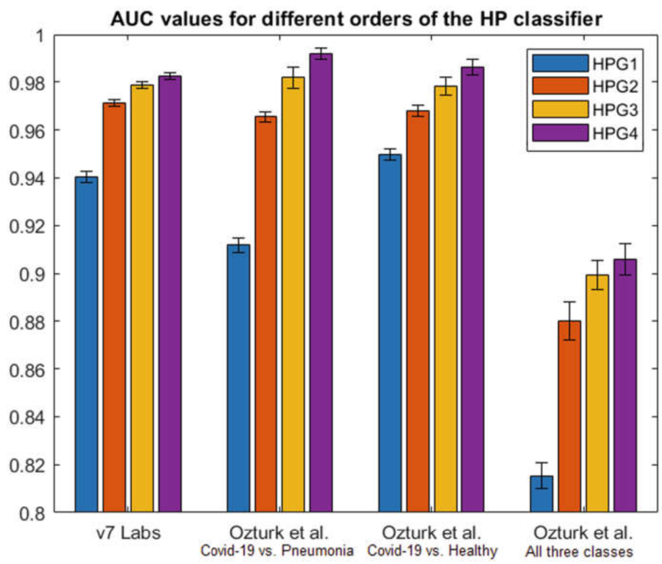


Fig. 5. Evaluation of the HP classifier with different orders.

Similarly to the results presented in [15], the 4th order for the polynomial function has presented the best results in all scenarios. Therefore, we chose the HPG4 for the following experiments. The tested scenarios were: Covid-19 vs. Healthy (v7 Labs dataset); Covid-19 vs. Pneumonia, Covid-19 vs. Healthy, and Covid-19 vs. Healthy vs. Pneumonia (Ozturk *et al.* dataset). The results obtained with AUC and ACC metrics and the standard deviation (SD) values are shown in Table 2.

Table 2. Evaluation of the HPG4 classifier using features obtained from the datasets.

Dataset	Groups	Features	AUC	SD	ACC	SD
v7 Labs	Covid-19 vs. Healthy	Local	0.9795	0.0011	97.09	0.0024
		Global	0.9712	0.0020	95.80	0.0031
		All	**0.9826**	0.0013	**97.89**	0.0015
Ozturk et al.	Covid-19 vs. Healthy	Local	0.9803	0.0047	96.83	0.0050
		Global	0.9734	0.0054	94.62	0.0055
		All	**0.9861**	0.0033	**97.93**	0.0036
	Covid-19 vs. Pneumonia	Local	0.9869	0.0026	97.86	0.0042
		Global	0.9718	0.0042	94.64	0.0067
		All	**0.9917**	0.0023	**98.37**	0.0042
	All classes	Local	0.8929	0.0093	91.31	0.0061
		Global	0.8674	0.0051	89.36	0.0044
		All	**0.9057**	0.0065	**92.13**	0.0048

The HP classifier in the 4th order, which will be indicated as HPG4 from now, has provided AUC values above 0.97 in all binary classification scenarios, considering both datasets. Nevertheless, these results have also shown that the combined feature vector containing both local and global percolation features has a better performance than using the features separately, as the highest AUC and accuracy rates occurred in this situation. Moreover, it can be noted that the Covid-19 vs Pneumonia classification has achieved better results than the comparison with the Healthy class in both datasets. As for the multiclass classification, an AUC value of 0.906 was obtained when use the full feature set, which is a value slightly higher than the one obtained for classifying global and local features separately. In the following tests, we compared the performance of the HPG4 with other five classifiers under these same scenarios. The results are shown in Table 3.

The classifications with the v7 Labs Dataset were the first to be evaluated. Although all of the six classifiers provided AUC values above 0.96, the HPG4, RoF and RaF algorithms are noteworthy for obtaining the best AUC and ACC values. The highest accuracy was provided by the HPG4 classifier using the

Table 3. Comparison among HPG4 and other five classifier regarding the classification of the combined feature set.

Dataset	Groups	Classifier	AUC	SD	ACC	SD
v7 Labs	Covid-19 vs. Healthy	HPG4	0.9826	0.0013	**97.89**	0.0015
		LGT	0.9777	0.0016	96.23	0.0024
		MLP	0.9832	0.0020	95.71	0.0046
		DEC	0.9834	0.0019	95.38	0.0036
		RoF	0.9884	0.0012	96.35	0.0034
		RaF	**0.9891**	0.0007	95.84	0.0013
Ozturk et al.	Covid-19 vs. Pneumonia	HPG4	**0.9917**	0.0023	**98.37**	0.0042
		LGT	0.9760	0.0043	94.43	0.0049
		MLP	0.9660	0.0040	92.70	0.0092
		DEC	0.9403	0.0097	91.48	0.0077
		RoF	0.9720	0.0055	94.78	0.0040
		RaF	0.9653	0.0036	93.00	0.0046
	Covid-19 vs. Healthy	HPG4	**0.9861**	0.0033	**97.93**	0.0036
		LGT	0.9746	0.0071	95.43	0.0040
		MLP	0.9581	0.0069	93.79	0.0078
		DEC	0.9689	0.0074	95.67	0.0063
		RoF	0.9770	0.0074	95.67	0.0063
		RaF	0.9787	0.0024	96.18	0.0026
	All classes	HPG4	**0.9057**	0.0065	**92.13**	0.0049
		LGT	0.8790	0.0058	89.68	0.0028
		MLP	0.8275	0.0102	86.42	0.0088
		DEC	0.8376	0.0089	87.03	0.0090
		RoF	0.8669	0.0056	88.84	0.0046
		RaF	0.8440	0.0068	87.65	0.0061

combination of global and local features and consisted in an ACC value of 97.89. The highest AUC value, which was 0.9891, was obtained by the RaF classifier.

The same experiments were also performed on the Ozturk *et al.* dataset. Three situations were evaluated: Covid-19 vs. Pneumonia classification; Covid-19 vs. Healthy classification; and Covid-19 vs. Pneumonia vs. Healthy classification. For the Covid-19 vs. Pneumonia comparison, the best results were also obtained using the HPG4 classifier, although the RaF algorithm did not perform as well as in the previous evaluated dataset, despite still having the second best performance in terms of AUC and ACC out of the six evaluated classifiers. In the Covid-19 vs. Healthy comparison, the HPG4 classifier has also provided the best AUC and ACC values, overcoming the obtained accuracy for the same class comparison in the v7 Labs dataset, which not only contains a greater number of samples, but these are also segmented.

The AUC and ACC values for the classification of the Covid-19 and Healthy groups were slightly smaller than the ones obtained for the Covid-19 vs. Pneumonia classification. However, we believe these results are influenced by the dataset composition, since the images for healthy and pneumonia samples and images for Covid-19 samples come from different sources, since other papers have reported high accuracy values for both comparisons [27,28]. Finally, for multiclass classification, the HPG4 classifier provided the best AUC and ACC results by a margin of at least 0.02 regarding the AUC. This may indicate a better adaptability for multiclass classification than the other five tested classifiers, although further analysis with different datasets and features are required.

We have also applied the non-parametric Friedman test in order to verify the average performance of each of the six tested classification algorithms considering the 48 features as input based on the AUC values. The classifiers that provided the highest AUC averages were the HPG4 (0.9665) and RoF (0.9511). The p-values for all pair-wise comparisons (Conover) are presented on Table 4, wherein statistically significant differences ($\alpha < 0.05$) are highlighted in bold. Although most of the *p-values* do not indicate a clear significant statistical difference among the classifiers, the best average ranking was obtained for HPG4, which reinforces its suitability for the classification of this type of features for both datasets.

Table 4. *p-values* obtained from the application of the Friedman test for comparing the performance of the six classifiers.

p-values	HPG4	LGT	MLP	DEC	RoF	RaF	Avg. Rank
HPG4	–	0.2268	**0.0235**	**0.0355**	0.5381	0.4140	2.0
LGT	0.2268	–	0.2268	0.3102	0.5381	0.6804	3.5
MLP	**0.0235**	0.2268	–	0.8365	0.0782	0.1136	5.0
DEC	**0.0355**	0.3102	0.8365	–	0.1136	0.1621	4.75
RoF	0.5381	0.5381	0.0782	0.1136	–	0.8365	2.75
RaF	0.4140	0.6804	0.1136	0.1621	0.8365	–	3.0

Overall, both local and global percolation features have shown to be a good discriminant factor in the classification of chest X-ray images regarding the detection of Covid-19, providing AUC values above 0.95 in all tested binary scenarios, considering both datasets. For the multiclass classification, the HPG4 classifier was the only one able to obtain AUC and ACC values above 0.9. Therefore, the HPG4 and RoF classifiers appear to be more suitable for these types of features, since the highest AUC and ACC values were obtained from them.

On Table 5, an overview of some of the published computer vision methods aimed at the detection of Covid-19 in chest X-ray images is presented. Our proposed method is on par with the other approaches available in the literature and we believe that percolation features can provide a relevant contribution to upcoming research regarding classification of chest X-ray images. Moreover, our approach is based on the extraction of handcrafted features and does not require the application of deep learning algorithms, thus generating a smaller computational demand than CNN methods, also due to considerably smaller dimension of the final feature vector.

Table 5. Overview of different approaches for detecting Covid-19 in chest X-ray images.

Method	Approach	Classes	Segmented	Samples	Features	ACC
[9]	DenseNet201	2	No	50	1,920	90.00%
[20]	DarkCovidNet	2	No	625	338	98.08%
		3		1,125		87.02%
[27]	Fuzzy color, MobileNetV2, SqueezeNet	3	No	458	1,357	99.27%
[28]	LBP, ReliefF, SVM	2	No	234	1,459	99.55%
[12]	Xception	3	No	6,432	2,048	97.97%
[11]	Shearlet transform, Entropy and Energy, ELM	2	No	561	98	99.29%
[7]	Haralick, Zernike features, SVM	4	No	6309	168	89.78%
[16]	GLCM, SVM ensembles	2	Yes	1,710	90	97.04%
[19]	Robust features, SAE, PCA, SVM	6	No	260	20	94.23%
[22]	Modified DarkCovidNet	2	No	4,572	507	99.53%
		3		10,190		94.18%
Proposed	Percolation features, HPG4	2	Yes	2,002	48	97.89%
		2	No	625		98.37%
		2		625		97.93%
		3		1,125		92.13%

4 Conclusion

In this paper, we proposed the use of local and global percolation features to represent chest X-ray images on the task of detecting infections by Covid-19. The method is based on [23] and was adapted to greyscale images. The HPG4 classifier was able to provide the highest ACC values on both tested datasets. We also verified that while global features were more significant for classifying images from the v7 Labs dataset, there was a more balanced contribution between the global and local features sets in classifying images from the Ozturk *et al.* dataset.

These obtained results are compatible with other methods that have been published since the beginning of the pandemic. We have shown that percolation based features can differentiate among Covid-19, Pneumonia and Healthy chest X-ray samples and encourage its application on different datasets and classification problems.

For future works, we suggest the application of this approach on new chest X-ray datasets containing more Covid-19 samples as well as the evaluation of these features on images from CT-scans. Other fractal measures such as lacunarity and fractal dimension could be evaluated as indicators of the presence of Covid-19 in chest X-ray images. Different parameters to obtain the features such as the maximum value of L and the type of distance could also be tested in order to improve the obtained results.

Funding. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, National Council for Scientific and Technological Development CNPq (Grants #132940/2019-1, #313643/2021-0 and #311404/2021-9), the State of Minas Gerais Research Foundation - FAPEMIG (Grants #APQ-00578-18 and #APQ-01129-21), the State of São Paulo Research Foundation - FAPESP (Grant #2022/03020-1). and the project NextGenAI - Center for Responsible AI (2022-C05i0102-02), supported by IAPMEI, and also by FCT plurianual funding for 2020-2023 of LIACC (UIDB/00027/2020 UIDP/00027/2020).

References

1. Bandirali, M., et al.: Chest radiograph findings in asymptomatic and minimally symptomatic quarantined patients in codogno, italy during covid-19 pandemic. Radiology **295**(3), E7–E7 (2020)
2. Căliman, A., Ivanovici, M.: Psoriasis image analysis using color lacunarity. In: 2012 13th International Conference on Optimization of Electrical and Electronic Equipment (OPTIM), pp. 1401–1406. IEEE (2012)
3. Chihara, T.S.: An introduction to orthogonal polynomials. Courier Corporation (2011)
4. Corman, V.M., et al.: Detection of 2019 novel coronavirus (2019-ncov) by real-time rt-pcr. Eurosurveillance **25**(3), 2000045 (2020)
5. Cozzi, D., et al.: Chest x-ray in new coronavirus disease 2019 (covid-19) infection: findings and correlation with clinical outcome. Radiol. Med. (Torino) **125**, 730–737 (2020)
6. Dong, E., Du, H., Gardner, L.: An interactive web-based dashboard to track covid-19 in real time. Lancet. Infect. Dis **20**(5), 533–534 (2020)
7. Gomes, J.C., et al.: Ikonos: an intelligent tool to support diagnosis of covid-19 by texture analysis of x-ray images. Res. Biomed. Eng. 1–14 (2020)
8. Hassantabar, S., Ahmadi, M., Sharifi, A.: Diagnosis and detection of infected tissue of covid-19 patients based on lung x-ray image using convolutional neural network approaches. Chaos, Solitons Fractals **140**, 110170 (2020)
9. Hemdan, E.E.D., Shouman, M.A., Karar, M.E.: Covidx-net: a framework of deep learning classifiers to diagnose covid-19 in x-ray images. arXiv preprint arXiv:2003.11055 (2020)
10. Hooshmand Moghaddam, V., Hamidzadeh, J.: New hermite orthogonal polynomial kernel and combined kernels in support vector machine classifier. Pattern Recogn. **60**, 921–935 (2016)
11. Ismael, A.M., Şengür, A.: The investigation of multiresolution approaches for chest x-ray image based covid-19 detection. Health Inform. Sci. Syst. **8**(1), 1–11 (2020)
12. Jain, R., Gupta, M., Taneja, S., Hemanth, D.J.: Deep learning based detection and analysis of covid-19 on chest x-ray images. Appl. Intell. **51**(3), 1690–1700 (2021)
13. Kurmi, Y., Chaurasia, V., Ganesh, N.: Tumor malignancy detection using histopathology imaging. J. Med. Imaging Radiation Sci. **50**(4), 514–528 (2019)

14. v7 Labs: Covid-19 x-ray dataset (2020). www.github.com/v7labs/covid-19-xray-dataset
15. Martins, A.S., et al.: A hermite polynomial algorithm for detection of lesions in lymphoma images. Pattern Anal. Appli. 1–13 (2020)
16. Mohammed, S., Alkinani, F., Hassan, Y.: Automatic computer aided diagnostic for covid-19 based on chest x-ray image and particle swarm intelligence. Inter. J. Intell. Eng. Syst. **13**(5), 63–73 (2020)
17. Nayak, S.R., Nayak, D.R., Sinha, U., Arora, V., Pachori, R.B.: Application of deep learning techniques for detection of covid-19 cases using chest x-ray images: a comprehensive study. Biomed. Signal Process. Control **64**, 102365 (2021)
18. Neves, L.A., et al.: Multi-scale lacunarity as an alternative to quantify and diagnose the behavior of prostate cancer. Expert Syst. Appl. **41**(11), 5017–5029 (2014)
19. Öztürk, Ş, Özkaya, U., Barstuğan, M.: Classification of coronavirus (covid-19) from x-ray and ct images using shrunken features. Int. J. Imaging Syst. Technol. **31**(1), 5–15 (2021)
20. Ozturk, T., Talo, M., Yildirim, E.A., Baloglu, U.B., Yildirim, O., Acharya, U.R.: Automated detection of covid-19 cases using deep neural networks with x-ray images. Comput. Biol. Med. **121**, 103792 (2020)
21. Padierna, L.C., Carpio, M., Rojas-Domínguez, A., Puga, H., Fraire, H.: A novel formulation of orthogonal polynomial kernel functions for svm classifiers: the gegenbauer family. Pattern Recogn. **84**, 211–225 (2018). https://doi.org/10.1016/j.patcog.2018.07.010, www.sciencedirect.com/science/article/pii/S0031320318302280
22. Redie, D.K., et al.: Diagnosis of covid-19 using chest x-ray images based on modified darkcovidnet model. Evolutionary Intell. 1–10 (2022)
23. Roberto, G.F., Nascimento, M.Z., Martins, A.S., Tosta, T.A., Faria, P.R., Neves, L.A.: Classification of breast and colorectal tumors based on percolation of color normalized images. Comput. Graph. **84**, 134–143 (2019)
24. Shanableh, T., Assaleh, K.: Feature modeling using polynomial classifiers and stepwise regression. Neurocomputing **73**(10–12), 1752–1759 (2010)
25. Sohrabi, C., et al.: World health organization declares global emergency: a review of the 2019 novel coronavirus (covid-19). Int. J. Surg. **76**, 71–76 (2020)
26. Thangavelu, S.: Hermite and laguerre semigroups: some recent developments. In: Seminaires et Congres (to appear) (2006)
27. Toğaçar, M., Ergen, B., Cömert, Z.: Covid-19 detection using deep learning models to exploit social mimic optimization and structured chest x-ray images using fuzzy color and stacking approaches. Comput. Biol. Med. **121**, 103805 (2020)
28. Tuncer, T., Dogan, S., Ozyurt, F.: An automated residual exemplar local binary pattern and iterative relieff based covid-19 detection method using chest x-ray image. Chemometr. Intell. Lab. Syst. **203**, 104054 (2020)
29. Velavan, T.P., Meyer, C.G.: The covid-19 epidemic. Tropical Med. Intern. Health **25**(3), 278 (2020)
30. Zanaty, E., Afifi, A.: Generalized hermite kernel function for support vector machine classifications. Int. J. Comput. Appl. **42**(8), 765–773 (2020)