REGULAR PAPER

A comparative study of linear type multiple instance learning techniques for detecting COVID-19 by chest X-ray images

Matteo Avolio[1,3] · Antonio Fuduli[1,3] · Eugenio Vocaturo[2,3] · Ester Zumpano[2,3]

Abstract

At the end of 2019, the World Health Organization (WHO) referred that the Public Health Commission of Hubei Province, China, reported cases of severe and unknown pneumonia. A new coronavirus, SARS-CoV-2, was identified as responsible for the lung infection, called COVID-19 (COronaVIrus Disease 2019). Although the definitive COVID-19 diagnosis is made through specific molecular tests, an early diagnosis by imaging became crucial to contain the spread, morbidity and mortality of the pandemic. In such context, chest X-ray radiography, as an element that assists the diagnosis allowing also the follow-up of the disease, plays a very important role since it is the most easily available and least expensive alternative. This work focuses on applying different linear type instance-level Multiple Instance Learning techniques to discriminate between COVID-19 and common viral pneumonia chest X-ray images, which is a difficult task due to the strong similarity characterizing the two classes. A relevant advantage of such approaches is that they are also suitable in terms of interpretability, as they easily allow clinicians to identify abnormal subregions in a positive radiographic image. Numerical experiments have been performed on a set of 200 images, obtaining the following results: accuracy = 95%, sensitivity = 99.29%, specificity = 91.24% and MCC = 0.9. The used algorithms appear promising in practical applications, taking into account their high speed and considering that no particular pre-processing techniques have been employed.

Keywords COVID-19 · X-ray image classification · Multiple instance learning

1 Introduction

Very recently the world has coped with the COVID-19 pandemic. COVID-19 is caused by a Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV-2) and its common

Matteo Avolio, Antonio Fuduli, Eugenio Vocaturo and Ester Zumpano have contributed equally to this work.

✉ Ester Zumpano
e.zumpano@dimes.unical.it

Matteo Avolio
matteo.avolio@unical.it

Antonio Fuduli
antonio.fuduli@unical.it

Eugenio Vocaturo
eugenio.vocaturo@cnr.it

[1] Department of Mathematics and Computer Science, University of Calabria, Rende, Italy

[2] DIMES, University of Calabria, Rende, Italy

[3] NANOTEC, National Research Council, Rende, Italy

symptoms are: fever, dry cough, fatigue, short breathing, vanishing of taste, loss of smell. The first known case of this Coronavirus disease was reported in Wuhan, China in the last days of 2019 [35] and, since then, the virus propagated all over the world. The main sources of infection are asymptomatic (but infected) people which can become a source of spread. Transmission mainly occurs by air through the *droplets*, but also by indirect transmission, such as through contact with contaminated surfaces. On March 11, 2020 the World Health Organization (WHO) declared the epidemic a global emergency (pandemic). Lockdown measures and drastic restrictions of movements and social life affected the lives of billions of people. COVID-19 was the most significant global crisis since the Second World War, but its repercussions exceeded those of a war.

COVID-19 has interstitial pneumonia as the predominant clinical manifestation. The interstitium is a particular entity located between the alveolus and the capillaries, which is investigated mainly with radiological techniques. Radiological imaging does not represent a diagnostic criterion for SARS-CoV2 infection, but it is able to highlight any pneu-

monia that can be associated with it, and in this case it is possible to see an opacity on the radiograph, called *thickening*. In the first four days (initial phase) the X-ray image is characterized by blurred thickening in the lower part of the lungs. From the fifth to the eighth day, there may be a clinical worsening of the patient who will present cough and difficulty in breathing (worsening dyspnea): in such case, the radiography image shows a greater extension of the pulmonary thickening and the lungs, so to speak, appear more and more white (Fig. 1).

Differently from computer tomography, X-ray imaging is cheaper and easier to perform: this simple and widely available imaging test can actually tell a lot about patient's clinical status and whether the COVID-19 patient requires hospitalization for mechanical ventilation or intubation. Finally a last, but non negligible, aspect is that the X-ray machines (even portable) are much more available also in poor and developing countries.

Table 1 summarizes the main clinical features which can be detected by any chest X-ray radiography of patients affected by COVID-19, reporting also the major pros and cons relating to its adoption.

As confirmed by [51], artificial intelligence systems nowadays play a very important role in supporting early diagnosis, illness evaluation and treatment response assessment for different diseases. As a consequence, we believe that using medical imaging techniques, in combination with more sophisticated machine learning systems, can effectively help also in the diagnosis and the follow-up of patients with COVID-19.

Since 2020 up today, several works have been published on COVID-19 detection by means of chest X-ray image classification. Most of them are convolutional neural network (CNN) approaches or, more in general, deep learning techniques. One of the first papers in this field is [27], where two algorithms were presented, including a deep neural network on fractal features of the images and CNN methods directly, the latter providing an accuracy of 93.2% in discriminating between COVID-19 and normal X-ray images.

In [4] an enhanced dense convolutional network (Dense Net) was proposed, while in [7] a new approach based on existing deep learning models was used, focusing on enhancing the pre-processing stage. In [37] the authors proposed a method based on an optimized robust CNN architecture, while in [45] a linear regression approach was designed in addition to a deep CNN, the former providing an accuracy of 97.6% for discrimination between healthy and COVID-19 patients. In [46], a combination of CNN, support vector machine (SVM) and Sobel filter was proposed on 1332 images, starting from 333 original images (77 images of COVID-19 patients and 256 images of normal subjects), whose number was increased to 1332 by a data augmentation operation. Three augmentation strategies (rotation, random

noise and horizontal flips) were also adopted in [26], while a fuzzy logic based on deep learning approach was proposed in [30] to differentiate between images of patients with COVID-19 pneumonia and with interstitial pneumonias not related to COVID-19, obtaining an accuracy of 80.9%. In [31], an evolutionary deep learning approach was designed to discriminate between COVID-19 and healthy patients, obtaining an accuracy of 98.57%. In [32] a stacked ensemble of four heterogeneous pre-trained computer vision deep learning models was presented, while in [38] a new deep learning framework was designed, based on the fusion of a dense convolutional network and a capsule neural network. In [17], a new pre-processing model for COVID-19 images was analyzed and a feature extraction was performed according to RGB values. In [42] a screening system was developed to differentiate among COVID-19, common viral pneumonia, bacterial pneumonia and normal chest X-ray images, using two different stages. The first stage is a pre-processing phase involving bone suppression and lung segmentation models, the former investigated also in [43]. The second stage is the classification task, based on a CNN system.

More recently, in [29] a deep learning method based on a custom CNN was proposed, utilizing dropout and batch normalization to enhance the performance and to reduce the overfitting. The proposed approach achieved a classification accuracy of 98.19% for discriminating among COVID-19, normal and common viral pneumonia images. In [48] an appropriate CNN model was enhanced using a feature fusion strategy from multi-modal imaging datasets and an SVM classifier was employed to discriminate between COVID-19 and healthy people, achieving an accuracy of 98.7%. In [22] after a pre-processing phase, the classification algorithm uses pre-trained CNN models, achieving an accuracy of 95% in the binary case between normal and COVID-19 X-ray images, while in [47] a densely attention mechanism-based network was proposed.

Other more general works on deep learning methods for COVID-19 X-ray classification are [33, 34, 49], while a survey on applying machine learning techniques in this field is reported in [19].

In this paper we present a comparative study of some instance-level Multiple Instance Learning (MIL) techniques applied to COVID-19 detection by means of chest X-ray images. As observed in the above literature review, most of the works in this field are deep learning models based on neural networks. Although the neural network approaches usually work well, on the other hand these models generally require a lot of data in the learning phase and, moreover, they are often uninterpretable, where by interpretability we mean the definition reported in [36], that is the degree to which an observer can understand the cause of a decision. Vice versa, as we will see in the next section, MIL approaches (especially the instance-level ones) are easily interpretable and
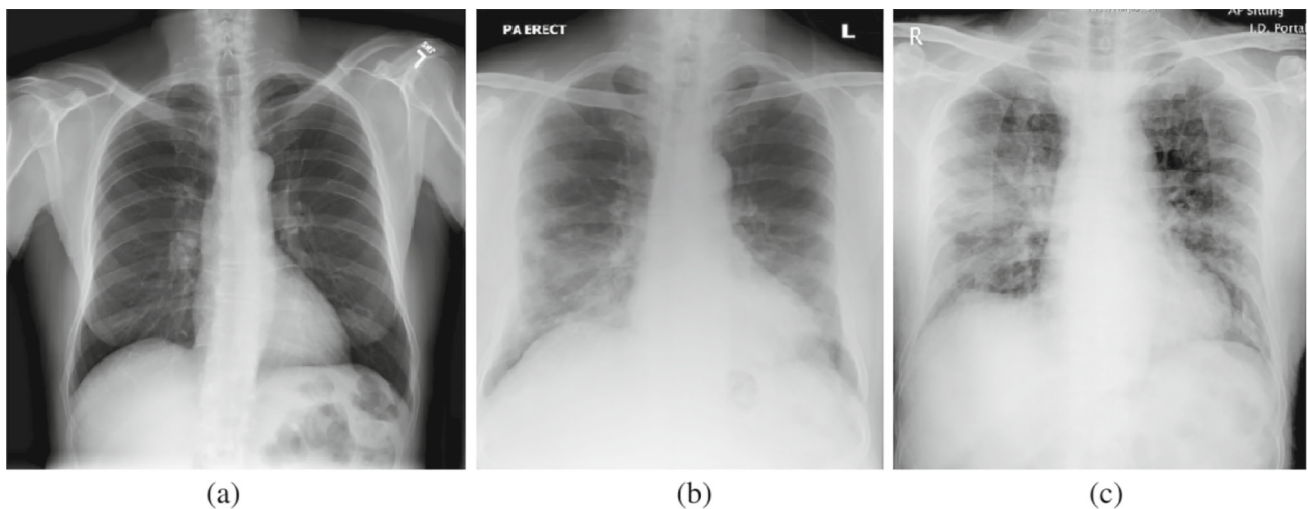
Fig. 1 Normal and COVID-19 X-ray images: **a** normal case, **b** COVID-19 at early stage, **c** COVID-19 at advanced stage [25]

Table 1 Clinical COVID-19 features of chest X-ray radiography: main advantages and disadvantages

Main clinical features	Advantages	Disadvantages
• Ground glass opacity (GGO) pulmonary nodules	• Easy to perform at the bed of the patient for follow-up	• Insufficient sensitivity to identify low density GGO
• Interstitial changes		
• Consolidation	• Availability	• Unable to detect pulmonary embolism and crazy paving
• Bilateral pneumonia		
• Post-inflammatory focal atelectatis		

consequently more suitable for image classification: in fact, once the discrimination task is performed, it is not difficult to practically interpret why an image is classified positive or negative. In addition, even though screening for COVID-19 no longer exists in most countries, we believe that this work is still relevant because it proposes to clinicians, if necessary, a new rapid and automatic approach to discriminate between different pneumonia.

The paper is organized in the following way. In the next section we recall the main concepts characterizing the MIL paradigm. In Sect. 3 we describe some recent linear type MIL techniques, that we have adopted in our numerical experimentation detailed in Sect. 4 and aimed at discriminating between COVID-19 and common viral pneumonia chest X-ray images. Finally some conclusions are drawn in Sect. 5.

2 Multiple instance learning

Multiple Instance Learning [28] (MIL) is a technique consisting in classifying sets of points: such sets are called *bags* and the points inside the sets are called *instances*. In com-

parison with the standard supervised classification, the main characteristic of a MIL approach is that in the learning phase only the class labels of the bags are known, whereas the class labels of the instances are unknown.

MIL techniques are applied in different contexts such as in bankruptcy prediction, image classification, text classification, speaker identification and so on. In particular, the first MIL problem encountered in the literature [21] concerned the classification of drug molecules (bags), on the basis of the possible three-dimensional conformations (instances) they can assume.

We focus on MIL classification with two classes of bags (positive and negative) and two classes of instances (positive and negative), using the so-called standard MIL assumption, which considers positive a bag containing at least a positive instance and negative a bag containing only negative instances. Such assumption fits very well diagnostic imaging (see [39]), where a patient is classified non-healthy (that is positive) if his/her medical image (bag) contains at least an abnormal subregion and is considered healthy (that is negative) if all the subregions forming his/her medical image are normal.

In the literature, for solving a MIL problem, there exist mainly three kinds of approaches (see [5, 18]). The first one is the bag-level approach where each bag is treated as a global entity, while the second one is the instance-level approach where the classification is performed in the instance space, obtaining the class label of each bag as aggregation of the class labels of the corresponding instances. The last approach is a compromise between the two previous ones: it consists in representing each bag by one of its instances, that will be successively used to perform the classification process.

Some recent MIL works in diagnostic imaging are [11, 15, 16, 50]. In particular, in [11] a MIL approach has been adopted for melanoma detection on clinical data constituted by some color dermoscopic images, with the aim of discriminating between melanomas (positive bags) and common nevi (negative bags) images. Extensive numerical experiments, performed on a dataset constituted by 80 melanomas and 80 common nevi images, provided an accuracy of 92.5%, with sensitivity and specificity equal to 97.5% and 87.5%, respectively. These results encourage to investigate on the possible use of MIL approaches also in COVID-19 detection by means of chest X-rays images, which is the scope of our study focusing on the discrimination between COVID-19 and common viral pneumonia patients, which is not an easy task due to the similarity of the X-ray images in the two classes.

The MIL techniques that we adopt to this end are described in the next subsection and fall into the instance-level class. All of them are designed to satisfy the standard MIL assumption, introduced above. As mentioned in the introduction section, the instance-level MIL techniques are really suitable in diagnostic imaging, especially in terms of interpretability: in fact an instance-level approach is aimed at assigning a label to each instance inside the bags, making a bag positive in case at least one of its instances is classified positive. Since in the MIL perspective the images are identified with the bags and the subregions forming the images are the corresponding instances, this criterion, based on the standard MIL assumption, allows the doctors to identify abnormal subregions in a positive X-ray image.

3 Some recent MIL linear type approaches

In this section we describe some recent MIL techniques, that we have used for our comparative study, aimed at discriminating between COVID-19 and viral pneumonia X-ray images. All these approaches provide a linear type separation of the bags, starting from the standard MIL assumption. In particular, while three of them (MIL-RL, mi-SPSVM and MIL-kink) provide a separation hyperplane, the last one (MI-POLY) separates the positive and the negative bags by constructing a polyhedron, based a prefixed finite number of hyperplanes.

We use the following notation. We denote by $u^T v$ the inner product between two n-dimensional real vectors $u, v \in \mathbb{R}^n$ and by $\|w\|$ the Euclidean norm of vector $w \in \mathbb{R}^n$. We indicate by m the number of positive bags, by k the number of negative bags and by $x_j \in \mathbb{R}^n$ a generic instance characterized by n features. Finally, we denote by J_i^+, for $i = 1, \ldots, m$, the index set corresponding to the instances of the i-th positive bag and by J_i^-, for $i = 1, \ldots, k$, the index set corresponding to the instances of the i-th negative bag.

3.1 MIL-RL

Algorithm MIL-RL [9] is an instance-level MIL technique, providing a separation hyperplane of the type:

$$H(w, b) \overset{\triangle}{=} \{x \in \mathbb{R}^n \mid w^T x + b = 0\}, \tag{1}$$

where $w \in \mathbb{R}^n$ is the normal to the hyperplane and $b \in \mathbb{R}$ is the bias. The algorithm is based on a heuristic solution to the following SVM type model introduced in [6]:

$$
\begin{cases}
\min_{y,w,b} \dfrac{1}{2}\|w\|^2 + C \displaystyle\sum_{i=1}^{m} \sum_{j \in J_i^+} \max\{0, 1 - y_j(w^T x_j + b)\} \\[2mm]
\qquad + C \displaystyle\sum_{i=1}^{k} \sum_{j \in J_i^-} \max\{0, 1 + (w^T x_j + b)\} \\[2mm]
\qquad \displaystyle\sum_{j \in J_i^+} \dfrac{y_j + 1}{2} \geq 1 \quad i = 1, \ldots, m \\[2mm]
\qquad y_j \in \{-1, +1\} \quad j \in J_i^+, \quad i = 1, \ldots, m,
\end{cases}
\tag{2}
$$

where the unknowns y_j represent the class labels to be assigned to the instances of the positive bags. As in the standard SVM approach, the positive parameter C tunes the weight between the maximization of the margin, obtained by minimizing the Euclidean norm of w, and the minimization of the misclassification errors of the instances, given by the second and the third term of the objective function. Finally, the constraints

$$\sum_{j \in J_i^+} \frac{y_j + 1}{2} \geq 1 \quad i = 1, \ldots, m \tag{3}$$

impose that, for each positive bag, at least one instance should be positive (that is with label equal to +1).

Note that, when $m = k = 1$ and $y_j = +1$ for any j, problem (2) reduces to the classical SVM quadratic program.

MIL-RL is based on solving successive Lagrangian relaxation problems of (2), obtained by relaxing constraints (3). In [9] it has been shown that, considering the Lagrangian dual of (2), in correspondence to the optimal solution there is no duality gap between the primal and dual objective functions.

3.2 mi-SPSVM

Algorithm mi-SPSVM, recently introduced in [13, 14], combines the nice properties exhibited for supervised classification by the SVM technique in terms of accuracy, and by the proximal support vector machine (PSVM) approach [24] in terms of computational complexity. It computes a separating hyperplane of the type (1) by solving, at each iteration, the following quadratic problem:

$$\min_{w,b} \frac{1}{2} \left\| \begin{matrix} w \\ b \end{matrix} \right\|^2 \frac{C}{2} \sum_{j \in J^+} [1 - (w^T x_j + b)]^2$$
$$+ C \sum_{j \in J^-} \max\{0, 1 + (w^T x_j + b)\}, \tag{4}$$

by varying of the sets J^+ and J^-, which contain the indexes of the instances currently considered positive and negative, respectively. In the separable case, while the third term of problem (4) makes the hyperplane

$$H^-(w, b) \overset{\triangle}{=} \{x \in \mathbb{R}^n \mid w^T x + b = -1\} \tag{5}$$

a supporting hyperplane for the instances indexed by J^- (as in the standard SVM), the second term maximizes the proximity of the positive instances (indexed by J^+) around the hyperplane

$$H^+(w, b) \overset{\triangle}{=} \{x \in \mathbb{R}^n \mid w^T x + b = 1\}. \tag{6}$$

Moreover, differently from SVM, in the PSVM approach the maximization of the margin is obtained including in the norm also the bias b, making problem (4) strictly convex also with respect to b.

In mi-SPSVM, the inizialization of the sets J^+ and J^- is done by inserting in J^+ the indexes of all the instances of the positive bags and in J^- the indexes of all the instances of the negative bags. Once an optimal solution (w^*, b^*) to problem (4) has been computed, the two index sets J^+ and J^- are updated in the following way:

$$J^+ := J^+ \setminus \bar{J} \quad \text{and} \quad J^- := J^- \cup \bar{J}$$

where

$$\bar{J} = \{j \in J^+ \setminus J^* \mid w^{*T} x_j + b^* \leq -1\},$$

with

$$J^* = \{j_i^*, \ i = 1, \dots, m \mid w^{*T} x_{j_i^*} + b^* \leq -1\}$$

and

$$j_i^* \overset{\triangle}{=} \arg \max_{j \in (J_i^+ \cap J^+)} \{w^{*T} x_j + b^*\}.$$

Note that a particular role in the definition of the set $\bar{J}$ is played by the set J^*, introduced for taking into account constraints (3), which impose the satisfaction of the standard MIL assumption. At the current iteration, the set J^* is the index set (subset of J^+) corresponding to the instances closest, for each positive bag, to the current hyperplane $H(w^*, b^*)$ and strictly lying in the negative side with respect to it. If an index, say $j_i^* \in J^*$, corresponding to one of such instances entered the set J^-, all the instances of the i-th positive bag would be considered negative by problem (4), favouring the violation of the standard MIL assumption. This is the reason why the indexes of J^* are prevented from entering the set J^-: in this way, for each positive bag, at least an index corresponding to one of its instances is guaranteed to be inside J^+.

3.3 MIL-kink

This approach, described in details in [23], provides a separation hyperplane of the type (1), by heuristically minimizing the following error function:

$$\sum_{i=1}^m \max \left\{ 0, \min_{j \in J_i^+} \{1 - (w^T x_j + b)\} \right\}$$
$$+ \sum_{i=1}^k \sum_{j \in J_i^-} \max \left\{ 0, 1 + (w^T x_j + b) \right\}, \tag{7}$$

which is easily derived by taking into account the standard MIL assumption. In fact a positive bag, indexed by J_i^+, is correctly classified if

$$w^T x_j + b \geq 1, \quad \text{for at least one } j \in J_i^+,$$

that is if

$$1 - (w^T x_j + b) \leq 0, \quad \text{for at least one } j \in J_i^+.$$

As a consequence the bag is misclassified if

$$1 - (w^T x_j + b) > 0, \quad \text{for each } j \in J_i^+,$$

that is if

$$\min_{j \in J_i^+} \{1 - (w^T x_j + b)\} > 0.$$

On the other hand, a negative bag, indexed by J_i^-, is correctly classified when

$$w^T x_j + b \leq -1, \quad \text{for each } j \in J_i^-,$$

that is when

$$\max_{j \in J_i^-} \{1 + (w^T x_j + b)\} \leq 0,$$

and, consequently, it is misclassified when

$$\max_{j \in J_i^-} \{1 + (w^T x_j + b)\} > 0.$$

Function (7) is very difficult to be minimized, since it is nonconvex and nonsmooth. For this reason, in [23], the authors proposed to adopt a very fast heuristic approach, based on computing the optimal value of b in correspondence to a prefixed value of w (judiciously chosen in advance) and in simply exploring the kink points of (7).

In [23] a variant of the algorithm was also proposed, on the basis of a simple modification of (7).

3.4 MI-POLY

Differently from the above MIL approaches, which provide a single separation hyperplane, Algorithm MI-POLY [8] is based on the concept of polyhedral separability, obtainable by generating a separating polyhedron by means of a finite number, say $h > 1$, of hyperplanes. In order to recall the MIL model proposed in [8], we first report the basic definition of polyhedral separability for supervised learning.

Let

$$\mathcal{P} = \{p_1, \dots, p_r\}, \quad \text{with } p_j \in \mathbb{R}^n, \ j = 1, \dots, r$$

and

$$\mathcal{Q} = \{q_1, \dots, q_s\}, \quad \text{with } q_j \in \mathbb{R}^n, \ j = 1, \dots, s,$$

be two disjoint point sets. They are polyhedrally separable [12] if and only if there exists a finite number h of hyperplanes

$$H_t(w_t, b_t) \stackrel{\triangle}{=} \{x \in \mathbb{R}^n \mid w_t^T x + b_t = 0\},$$

with $w_t \in \mathbb{R}^n$ and $b_t \in \mathbb{R}$, for $t = 1, \dots, h$, such that, for all $j = 1, \dots, r$,

$$w_t^T p_j + b_t \leq -1, \quad \text{for all } t = 1, \dots, h$$

and, for all $j = 1, \dots, s$,

$$w_t^T q_j + b_t \geq 1, \quad \text{for at least an index } t \in \{1, \dots, h\}.$$

On the basis of the above definition, the two sets $\mathcal{P}$ and $\mathcal{Q}$ are polyhedrally separable if there exists a polyhedron generated by a finite number h of hyperplane, such that all points of $\mathcal{P}$ are inside the polyhedron and all points of $\mathcal{Q}$ are outside.

To extend the supervised polyhedral separation to MIL, in [8] the authors proposed to heuristically solve the following nonsmooth nonconvex optimization problem, by means of DC (Difference of Convex functions) techniques:

$$\begin{cases} \min_{y,w,b} \dfrac{1}{2} \sum_{t=1}^{h} \|w_t\|^2 + C \sum_{i=1}^{m} \sum_{j \in J_i^+} \\ \quad + \max\{0, 1 + y_j \max_{1 \leq t \leq h} (w_t^T x_j + b_t)\} \\ +C \sum_{i=1}^{k} \sum_{j \in J_i^-} \max\{0, 1 - \max_{1 \leq t \leq h} (w_t^T x_j + b_t)\} \\ \sum_{j \in J_i^+} \dfrac{y_j + 1}{2} \geq 1 \quad i = 1, \dots, m \\ y_j \in \{-1, +1\} \quad j \in J_i^+, \quad i = 1, \dots, m \end{cases}$$

$$(8)$$

Problem (8) exploits the standard MIL assumption, by imposing that, for each positive bag, at least an instance should lie inside the polyhedron and, for each negative bag, all the instances should be outside. Similarly to model (2), the first term of the objective function is aimed at maximizing the margins in correspondence to the h hyperplanes, while the successive two terms minimize the misclassification error of the instances belonging to the positive and negative bags, respectively. Note that, in case $h = 1$, problem (8) reduces exactly to problem (2) taking into account the symmetric role played by the two halfspaces generated by a single hyperplane.

4 A comparative study on chest X-ray images

We have performed a comparative study of the MIL algorithms described in the previous section, with the aim at discriminating between COVID-19 (positive X-ray images) and common viral pneumonia (negative X-ray images). The flowchart of the overall experimentation is reported in Fig. 2.

All the algorithms have been run on a Microsoft Windows 11 system, characterized by 16 GB of RAM and a 2.30 GHz Intel Core i7 processor. They have been tested on 200 X-ray chest images, randomly taken from the public dataset described in Sect. 4.1: 100 images are relative to

Fig. 2 Flowchart of the experimentation: MIL-RL [9], mi-SPSVM [13] and MIL-kink [23] generate a separating hyperplane, while MI-POLY [8] a separating polyhedron. On the basis of the standard MIL assumption, the blue outlined images are classified as (positive) COVID-19 images, while the red outlined ones are classified as (negative) common viral pneumonia images (colour figure online)
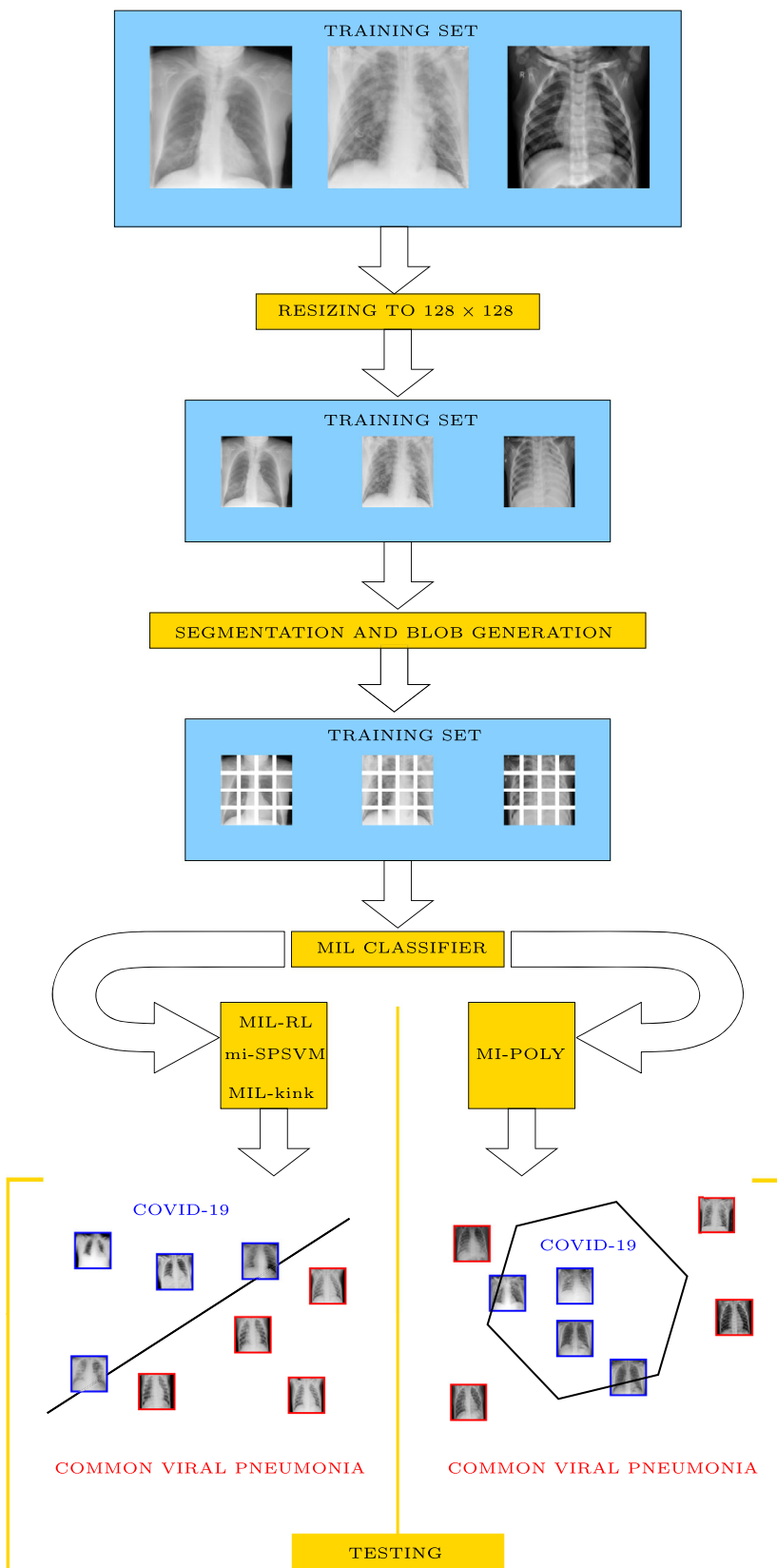
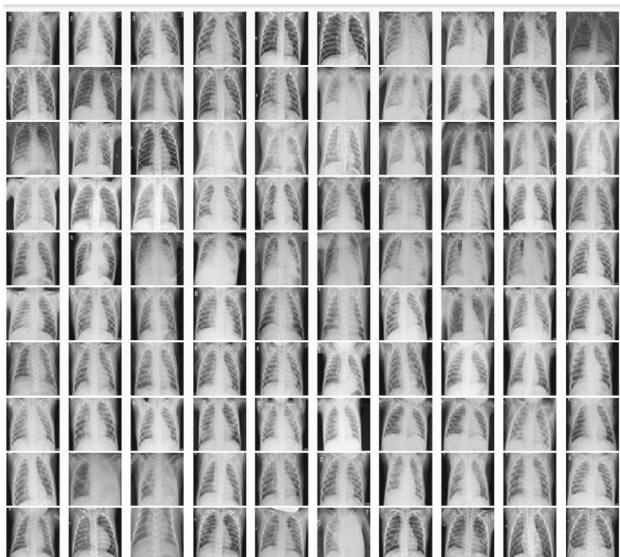**Fig. 3** 100 COVID-19 chest X-ray images (positive images)

Fig. 4 100 viral pneumonia chest X-ray images (negative images)

people affected by COVID-19 (Fig. 3) and 100 correspond to people with common viral pneumonia (Fig. 4).

4.1 Dataset description

The original dataset from which we have drawn the images used in our computational study is named COVID-19 Radiography Database [40] and it is detailed also in [20, 41]. The overall dataset is constituted by 3616 COVID-19 images, 10,192 normal images, 6012 lung opacity (that is non-COVID lung infection) images and 1345 common viral pneumonia images. Among the COVID-19 images, 2473 were collected from the BIMCV-COVID19+ dataset [1], 183

from a German medical school [2], 559 from the (SIRM), GitHub, Kaggle & Twitter, and 400 from another repository [3].

4.2 Implementation details, segmentation and futures

We have used the same MATLAB implementations of MIL-RL adopted in [11], of mi-SPSVM tested in [13] and of MI-POLY presented in [8]. About MIL-kink, we have used both the implementations tested in [23] and named MIL-kink[1] and MIL-kink[2], corresponding, respectively, to the minimization of function (7) and to its variant cited at the end of Sect. 3.3 when w is fixed. For each code, we have maintained the optimal tuning of the parameters described in the respective papers.

As for the segmentation process, we have adopted a procedure similar to that one used in [10] and [50]. In particular, we have reduced the resolution of each image to 128×128 pixels dimension and we have grouped the pixels in appropriate square subregions (blobs). In this way, each image is represented as a bag, while a blob corresponds to an instance of the bag. For each instance (blob), we have first considered the following 10 features:

- the average and the variance of the grey-scale intensity of the blob: 2 features;
- the differences between the average of the grey-scale intensity of the blob and that ones of the adjacent blobs (upper, lower, left, right): 4 features;
- the differences between the variance of the grey-scale intensity of the blob and that ones of the adjacent blobs (upper, lower, left, right): 4 features.

To exploit information about the texture of the images, for each blob we have also computed the corresponding grey-scale co-occurence matrix, by using the `graycomatrix` subroutine provided by the Image Processing MATLAB toolbox. In particular, fixing the number of gray levels equal to 3, for each blob we have generated a 3×3 co-occurence matrix, having in this way a total number of features equal to 19.

4.3 Numerical results

In order to consider different sizes of the testing and the training sets, we have used two validation protocols: the 5-fold cross-validation (5-CV) and the 10-fold cross-validation (10-CV). In Table 2 and in Table 3, we report the respective average results computed on the testing set and expressed in terms of the following performance indicators:

Table 2 Dataset constituted by 100 COVID-19 and 100 viral pneumonia chest X-ray images: 5-CV average testing values, provided by MIL-RL [9], mi-SPSVM [13], MIL-kink [23] and MI-POLY [8]

5-CV	MIL-RL	mi-SPSVM	MIL-kink[1]	MIL-kink[2]	MI-POLY
CPU time (s)	1.84	0.18	0.01	**0.00**	8.99
Accuracy (%)	93.50	**94.50**	78.50	80.00	92.50
Sensitivity (%)	98.26	98.08	75.36	78.46	**98.89**
Specificity (%)	89.29	**91.02**	81.43	81.43	86.38
PPV (%)	89.98	**91.57**	79.67	80.31	87.89
NPV (%)	97.78	97.75	77.65	79.84	**98.75**
F-score (%)	93.78	**94.67**	77.24	79.24	92.92
κ	0.87	**0.89**	0.57	0.60	0.85
MCC	0.88	**0.89**	0.57	0.60	0.86

Table 3 Dataset constituted by 100 COVID-19 and 100 viral pneumonia chest X-ray images: 10-CV average testing values, provided by MIL-RL [9], mi-SPSVM [13], MIL-kink [23] and MI-POLY [8]

10-CV	MIL-RL	mi-SPSVM	MIL-kink[1]	MIL-kink[2]	MI-POLY
CPU time (s)	2.31	0.38	**0.00**	**0.00**	13.32
Accuracy (%)	93.50	**95.00**	79.00	79.50	94.50
Sensitivity (%)	98.57	**99.29**	76.64	78.46	98.04
Specificity (%)	89.74	91.24	81.22	80.38	**91.52**
PPV (%)	89.72	**91.69**	79.40	78.82	91.27
NPV (%)	97.50	**98.57**	78.27	79.45	97.74
F-score (%)	93.64	**95.18**	77.51	78.27	94.37
κ	0.87	**0.90**	0.57	0.58	0.89
MCC	0.88	**0.90**	0.58	0.59	0.89

- $accuracy = \dfrac{TP+TN}{TP+TN+FP+FN} \in [0, 1]$: it provides the proportion of the correctly classified images, with respect to the overall dataset.
- $sensitivity = \dfrac{TP}{TP+FN} \in [0, 1]$: called also *true positive rate* or *recall*, it measures the proportion of the correctly classified positive images, with respect to the total number of positive images.
- $specificity = \dfrac{TN}{TN+FP} \in [0, 1]$: called also *true negative rate*, it measures the proportion of the correctly classified negative images, with respect to the total number of negative images.
- $PPV = \dfrac{TP}{TP+FP} \in [0, 1]$: called also *precision*, it is the *positive predictive value* and it measures the proportion of the correctly classified positive images, with respect to the total number of images classified as positive.
- $NPV = \dfrac{TN}{TN+FN} \in [0, 1]$: it is the *negative predictive value* and it measures the proportion of the correctly classified negative images, with respect to the total number of images classified as negative.
- $F\text{-}score = 2\dfrac{sensitivity \cdot PPV}{sensitivity + PPV} \in [0, 1]$: it is the harmonic mean of sensitivity and PPV.
- $\kappa = 2\cdot\dfrac{TP\cdot TN - FP\cdot FN}{(TP+FP)\cdot(FP+TN)+(TP+FN)\cdot(TN+FN)} \in [-1, 1]$: it is the *Cohen's kappa coefficient*, providing a measure

of the agreement between the actual and the predicted observations.
- $MCC = \dfrac{TP\cdot TN - FP\cdot FN}{\sqrt{(TP+FP)\cdot(TP+FN)\cdot(TN+FP)\cdot(TN+FN)}} \in [-1, 1]$: it is the *Matthews correlation coefficient*.

In the above list, the quantities TP, TN, FP and FN are the entries of the so-called 2×2 *confusion matrix*. In particular, TP (*true positive*) indicates the number of correctly classified positive images, TN (*true negative*) is the number of correctly classified negative images, FP (*false positive*) denotes the number of misclassified negative images and FN (*false negative*) is the number of misclassified positive images.

In Tables 2 and 3, we also report the average CPU time spent by the classifier to determine the optimal separation surface in the learning phase. For each of these performance parameters, the best result is highlighted in bold.

Looking at the the two tables, we observe that, apart from MIL-kink[1] and MIL-kink[2], the three codes MIL-RL, mi-SPSVM and MI-POLY provide quite comparable results (reaching in all the cases an accuracy greater than 92%), even if mi-SPSVM is the best performing also in terms of CPU time. On the other hand, both the versions of Algorithm MIL-kink have a non negligible advantage in terms of speed, providing however a reasonable accuracy close to 80%. MIL-RL, mi-SPSVM and MI-POLY performs well also in terms of MCC, which is a more informative parameter than F1-score,

since it takes into account the number of correctly classified negative images (TN), that is not considered by the F1-score indicator.

Finally it is worth noting that the Cohen's kappa coefficient κ is fully aligned with MCC and it shows an almost perfect agreement between the actual and the predicted observations in case of MIL-RL, mi-SPSVM and MI-POLY, while for MIL-kink it provides a moderate agreement.

5 Conclusions

COVID-19 quickly spread around the world creating an emergency situation. The use of X-ray chest radiographs is an element that assists the diagnosis, allowing also the follow-up of the disease.

In this context, we have focused on Multiple Instance Learning (MIL) techniques, which have proven to be effective in image analysis. In particular, we have presented a comparative study of some very recent MIL approaches, tested on a set of 200 chest X-ray images with the aim of discriminating between COVID-19 and common viral pneumonia. This study has highlighted the great potentiality of MIL, providing an accuracy result equal to 95% and a value of MCC equal to 0.9. Differently from the deep learning techniques, MIL approaches (especially the instance-level ones) are easily interpretable and consequently more suitable for image classification, since it is not difficult to practically interpret why an image is classified positive or negative.

The promising results of this study open up several avenues for future research, which may involve the implementation of more complex MIL systems, including additional features and more sophisticated segmentation and pre-processing techniques, such as those used in [44]. Combination of MIL with deep learning approaches could exploit the strengths of both methodologies. For instance, hybrid models that use MIL for interpretability and CNNs for feature extraction might offer higher performance. A further direction of research aims to include a wider range of cases, encompassing various stages of COVID-19 and other respiratory conditions, across different populations and healthcare settings.

Data and software code availability The data and the codes utilized during the current study are available in the following repository: https://github.com/antfud/MIL-COVID19.

References

1. BIMCV-COVID19+: a large annotated dataset of RX and CT images of COVID-19 patients. https://bimcv.cipf.es/bimcv-projects/bimcv-covid19/#1590858128006-9e640421-6711
2. COVID-19-image-repository. https://github.com/ml-workgroup/covid-19-image-repository/tree/master/png
3. COVID-CXNet. https://github.com/armiro/COVID-CXNet
4. Albahli, S., Ayub, N., Shiraz, M.: Coronavirus disease (COVID-19) detection using X-ray images and enhanced DenseNet. Appl. Soft Comput. **110**, 107645 (2021)
5. Amores, J.: Multiple instance classification: review, taxonomy and comparative study. Artif. Intell. **201**, 81–105 (2013)
6. Andrews, S., Tsochantaridis, I., Hofmann, T.: Support vector machines for multiple-instance learning. In: Becker, S., Thrun, S., Obermayer, K. (eds.) Advances in Neural Information Processing Systems, pp. 561–568. MIT Press, Cambridge (2003)
7. Arias-Garzón, D., Alzate-Grisales, J.A., Orozco-Arias, S., Arteaga-Arteaga, H.B., Bravo-Ortiz, M.A., Mora-Rubio, A., Saborit-Torres, J.M., Ángel Montell Serrano, J., de la Iglesia Vayá, M., Cardona-Morales, O., Tabares-Soto, R.: COVID-19 detection in X-ray images using convolutional neural networks. *Mach. Learn. Appl.*, **6**, 100138 (2021)
8. Astorino, A., Avolio, M., Fuduli, A.: Maximum-margin polyhedral separation for binary multiple instance learning. EURO J. Comput. Optim. **11**, 100070 (2023)
9. Astorino, A., Fuduli, A., Gaudioso, M.: A Lagrangian relaxation approach for binary multiple instance classification. IEEE Trans. Neural Netw. Learn. Syst. **30**(9), 2662–2671 (2019)
10. Astorino, A., Fuduli, A., Gaudioso, M., Vocaturo, E.: Multiple instance learning algorithm for medical image classification. In: Mecella, M., Amato, G., Gennaro, C., (eds) Proceedings of the 27th Italian Symposium on Advanced Database Systems, Castiglione della Pescaia (Grosseto), Italy, June 16–19, 2019, vol. 2400 of CEUR Workshop Proceedings (2019)
11. Astorino, A., Fuduli, A., Veltri, P., Vocaturo, E.: Melanoma detection by means of multiple instance learning. Interdiscip. Sci.: Comput. Life Sci. **12**(1), 24–31 (2020)
12. Astorino, A., Gaudioso, M.: Polyhedral separability through successive LP. J. Optim. Theory Appl. **112**(2), 265–293 (2002)
13. Avolio, M., Fuduli, A.: A semiproximal support vector machine approach for binary multiple instance learning. IEEE Trans. Neural Netw. Learn. Syst. **32**(8), 3566–3577 (2021)
14. Avolio, M., Fuduli, A.: The semiproximal SVM approach for multiple instance learning: a kernel-based computational study. Optim. Lett. **18**(2), 635–649 (2024)
15. Avolio, M., Fuduli, A., Vocaturo, E., Zumpano, E.: Multiple instance learning for viral pneumonia chest X-ray classification. In: CEUR Workshop Proceedings, vol. 3194, pp. 359–366 (2022)
16. Avolio, M., Fuduli, A., Vocaturo, E., Zumpano, E.: On detection of diabetic retinopathy via multiple instance learning. In: ACM International Conference Proceeding Series, pp. 170–176, (2023)
17. Avuçlu, E.: COVID-19 detection using X-ray images and statistical measurements. Measurement **201**, 111702 (2022)

18. Carbonneau, M., Cheplygina, V., Granger, E., Gagnon, G.: Multiple instance learning: a survey of problem characteristics and applications. Pattern Recognit. **77**, 329–353 (2018)

19. Caroprese, L., Vocaturo, E., Zumpano, E.: Machine learning techniques on X-ray images for COVID-19 classification. In: 2022 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), pp. 539–543 (2022)

20. Chowdhury, M.E.H., Rahman, T., Khandakar, A., Mazhar, R., Kadir, M.A., Mahbub, Z.B., Islam, K.R., Khan, M.S., Iqbal, A., Emadi, N.A., Reaz, M.B.I., Islam, M.T.: Can AI help in screening viral and COVID-19 pneumonia? IEEE Access **8**, 132665–132676 (2020)

21. Dietterich, T.G., Lathrop, R.H., Lozano-Pérez, T.: Solving the multiple instance problem with axis-parallel rectangles. Artif. Intell. **89**(1–2), 31–71 (1997)

22. El Houby, E.M.F.: COVID-19 detection from chest X-ray images using transfer learning. *Sci. Rep.*, **14**(1) (2024)

23. Fuduli, A., Gaudioso, M., Khalaf, W., Vocaturo, E.: A heuristic approach for multiple instance learning by linear separation. Soft. Comput. **26**(7), 3361–3368 (2022)

24. Fung, G. Mangasarian, O.: Proximal support vector machine classifiers. In: Provost, F., Srikant, R., (eds.) Proceedings KDD-2001: Knowledge Discovery and Data Mining, pp. 77–86. ACM (2001)

25. Gil, D., Díaz-Chito, K., Sánchez, C., Hernández-Sabaté, A.: Early screening of SARS-CoV-2 by intelligent analysis of X-ray images (2020)

26. Guefrechi, S., Jabra, M.B., Ammar, A., Koubaa, A., Hamam, H.: Deep learning based detection of COVID-19 from chest X-ray images. Multimedia Tools Appl. **80**(21–23), 31803–31820 (2021)

27. 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)

28. Herrera, F., Ventura, S., Bello, R., Cornelis, C., Zafra, A., Sánchez-Tarragó, D., Vluymans, S.: Multiple Instance Learning: Foundations and Algorithms. Springer, Cham (2016)

29. Hussein, A.M., Sharifai, A.G., Alia, O.M., Abualigah, L., Almotairi, K.H., Abujayyab, S.K.M., Gandomi, A.H.: Auto-detection of the coronavirus disease by using deep convolutional neural networks and X-ray photographs. Sci. Rep., **14**(1) (2024)

30. Ieracitano, C., Mammone, N., Versaci, M., Varone, G., Ali, A.-R., Armentano, A., Calabrese, G., Ferrarelli, A., Turano, L., Tebala, C., Hussain, Z., Sheikh, Z., Sheikh, A., Sceni, G., Hussain, A., Morabito, F.C.: A fuzzy-enhanced deep learning approach for early detection of COVID-19 pneumonia from portable chest X-ray images. Neurocomputing **481**, 202–215 (2022)

31. Jalali, S.M.J., Ahmadian, M., Ahmadian, S., Hedjam, R., Khosravi, A., Nahavandi, S.: X-ray image based COVID-19 detection using evolutionary deep learning approach. Expert Syst. Appl. **201**, 116942 (2022)

32. Jangam, E., Barreto, A.A.D., Annavarapu, C.S.R.: Automatic detection of COVID-19 from chest ct scan and chest X-Rays images using deep learning, transfer learning and stacking. Appl. Intell. **52**(2), 2243–2259 (2022)

33. Khero, K., Usman, M., Fong, A.: Deep learning framework for early detection of COVID-19 using X-ray images. Multimedia Tools Appl. **83**(3), 6883–6908 (2024)

34. Koyyada, S.P, Singh, T.P.: A systematic survey of automatic detection of lung diseases from chest X-ray images: COVID-19, pneumonia, and tuberculosis. SN Comput. Sci., **5**(2) (2024)

35. Li, Q., Guan, X., Wu, P., et al.: Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N. Engl. J. Med. **382**(13), 1199–1207 (2020)

36. Miller, T.: Explanation in artificial intelligence: insights from the social sciences. Artif. Intell. **267**, 1–38 (2019)

37. Pathan, S., Siddalingaswamy, P., Ali, T.: Automated detection of COVID-19 from chest X-ray scans using an optimized CNN architecture. Appl. Soft Comput. **104**, 107238 (2021)

38. Quan, H., Xu, X., Zheng, T., Li, Z., Zhao, M., Cui, X.: Densecapsnet: detection of COVID-19 from X-ray images using a capsule neural network. Comput. Biol. Med. **133**, 104399 (2021)

39. Quellec, G., Cazuguel, G., Cochener, B., Lamard, M.: Multiple-instance learning for medical image and video analysis. IEEE Rev. Biomed. Eng. **10**, 213–234 (2017)

40. Rahman, T., Chowdhury, M., Khandakar, A.: COVID-19 radiography database. https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database

41. Rahman, T., Khandakar, A., Qiblawey, Y., Tahir, A., Kiranyaz, S., Abul Kashem, S.B., Islam, M.T., Al Maadeed, S., Zughaier, S.M., Khan, M.S., Chowdhury, M.E.: Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images. Comput. Biol. Med. **132**, 104319 (2021)

42. Rani, G., Misra, A., Dhaka, V.S., Buddhi, D., Sharma, R.K., Zumpano, E., Vocaturo, E.: A multi-modal bone suppression, lung segmentation, and classification approach for accurate COVID-19 detection using chest radiographs. Intell. Syst. Appl. **16**, 200148 (2022)

43. Rani, G., Misra, A., Dhaka, V.S., Zumpano, E., Vocaturo, E.: Spatial feature and resolution maximization gan for bone suppression in chest radiographs. Comput. Methods Programs Biomed. **224**, 107024 (2022)

44. Rani, G., Thakkar, P., Verma, A., Mehta, V., Chavan, R., Dhaka, V.S., Sharma, R.K., Vocaturo, E., Zumpano, E.: KUB-UNet: segmentation of organs of urinary system from a KUB X-ray image. Comput. Methods Programs Biomed. **224**, 107031 (2022)

45. Rasheed, J., Hameed, A.A., Djeddi, C., Jamil, A., Al-Turjman, F.: A machine learning-based framework for diagnosis of COVID-19 from chest x-ray images. Interdiscip. Sci. - Comput. Life Sci. **13**(1), 103–117 (2021)

46. Sharifrazi, D., Alizadehsani, R., Roshanzamir, M., Joloudari, J.H., Shoeibi, A., Jafari, M., Hussain, S., Sani, Z.A., Hasanzadeh, F., Khozeimeh, F., Khosravi, A., Nahavandi, S., Panahiazar, M., Zare, A., Islam, S.M.S., Acharya, U.R.: Fusion of convolution neural network, support vector machine and sobel filter for accurate detection of COVID-19 patients using X-ray images. Biomed. Signal Process. Control **68**, 102622 (2021)

47. Ullah, Z., Usman, M., Latif, S., Gwak, J.: Densely attention mechanism based network for COVID-19 detection in chest X-rays. Sci. Rep. **13**(1) (2023)

48. Veluchamy, S., Sudharson, S., Annamalai, R., Bassfar, Z., Aljaedi, A., Jamal, S.: Automated detection of COVID? 19 from multimodal imaging data using optimized convolutional neural network model. J. Imaging Inform. Med., (in press) (2024)

49. Vocaturo, E., Zumpano,E., Caroprese, L.: Convolutional neural network techniques on X-ray images for COVID-19 classification. In: 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 3113–3115 (2021)

50. Zumpano, E., Fuduli, A., Vocaturo, E., Avolio, M.: Viral pneumonia images classification by multiple instance learning: preliminary results. In: ACM International Conference Proceeding Series, pp. 292–296 (2021)

51. Zumpano, E., Iaquinta, P., Caroprese, L., Dattola, F., Tradigo, G., Veltri, P., Vocaturo, E.: SIMPATICO 3D mobile for diagnostic procedures. In: ACM International Conference Proceeding Series (2019)