arXiv:2010.08115v2 [eess.IV] 5 Jun 2021

Pinball-OCSVM for early-stage COVID-19 diagnosis with limited posteroanterior chest X-ray images

Sanjay Kumar Sonbhadra

IIIT Allahabad, Prayagraj, U. P. India
rsi2017502@iiita.ac.in

Sonali Agarwal

IIIT Allahabad, Prayagraj, U. P. India
sonali@iiita.ac.in

P. Nagabhushan

IIIT Allahabad, Prayagraj, U. P. India
pnagabhushan@iiita.ac.in

The infection of respiratory coronavirus disease 2019 (COVID-19) starts with the upper respiratory tract and as the virus grows, the infection can progress to lungs and develop pneumonia. The conventional way of COVID-19 diagnosis is reverse transcription polymerase chain reaction (RT-PCR), which is less sensitive during early stages; especially if the patient is asymptomatic, which may further cause more severe pneumonia. In this context, several deep learning models have been proposed to identify pulmonary infections using publicly available chest X-ray (CXR) image datasets for early diagnosis, better treatment and quick cure. In these datasets, presence of less number of COVID-19 positive samples compared to other classes (normal, pneumonia and Tuberculosis) raises the challenge for unbiased learning of deep learning models. All deep learning models opted class balancing techniques to solve this issue; which however should be avoided in any medical diagnosis process. Moreover, the deep learning models are also data hungry and need massive computation resources. Therefore for quicker diagnosis, this research proposes a novel pinball loss function based one-class support vector machine (PB-OCSVM), that can work in presence of limited COVID-19 positive CXR samples with objectives to maximize the learning efficiency and to minimize the false predictions. The performance of the proposed model is compared with conventional OCSVM and existing deep learning models, and the experimental results prove that the proposed model outperformed over state-of-the-art methods. To validate the robustness of the proposed model, experiments are also performed with noisy CXR images and UCI benchmark datasets.

Keywords: COVID-19; Chest X-ray; Pneumonia; Classification; Deep learning; Pinball loss; One-class support vector machine.

1. Introduction

The respiratory coronavirus disease 2019 (COVID-19) is caused by severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) that is the most recently identified

2   *S. K. Sonbhadra et al.*

member of the coronavirus family [36]. This deadly disease was initially reported during late December, 2020 and spread to all the countries worldwide [38]. World health organization (WHO) declared this infectious disease as a public health emergency of international concern (PHEIC) on January 30, 2020 as it reached to many countries [43] and on Feb 11, 2020 named it "COVID-19". On March 11, 2020 WHO declared this a pandemic [44]. After the fourteen months journey, this virus caused over 160.5 million infections and more than 3.3 million deaths worldwide till May 12, 2021 [16]. This fatal disease is highly infectious and to control its spread, following three ways have been suggested as most promising preventive measures: social distancing [32], use of mask [8] and early identification of infected people. The social distancing and mask are the established preventive measures that people are strictly following, but early diagnosis is still a challenge to the research community. COVID-19 is highly infectious, thus the active patient must be quarantined to break the chain of infection. The best known method of COVID-19 diagnosis is real-time reverse transcription-polymerase chain reaction (RT-PCR) that came into effect after the rapid antibody tests showed unreliable results, because antibodies appear after 9-28 days of the infection and by this time, an infected person may spread the disease, if not isolated [7]. The RT-PCR test is a costly and time taking process, whereas inability to early-stage diagnosis and due to rapid growth of infections, medical experts are continuously trying to search some other ways of diagnosis.

SARS-CoV-2 virus initially affects the respiratory system of the infected person and in later stages, it affects lungs that may cause severe pneumonia [12]. Statistics show that ∼14% of the COVID-19 patients have shortness of breath and severe cough due to pneumonia, because as the viral infection increases, it damages the alveoli (small air sacs) and surrounding tissues. It is evident that the symptoms of COVID-19 are extremely heterogeneous, ranging from minimal symptoms to significant hypoxia with acute respiratory distress syndrome (ARDS) [25]. Like flu, the common early symptoms of COVID-19 are fever, dry cough, nausea, diarrhea, muscle aches, vomiting, headache, loss of smell or taste, sore throat, etc. COVID-19 starts affecting the lungs as soon as it reaches through the nose or throat. Statistics show that 40%-45% of COVID-19 patients remain asymptomatic, but the virus affects their respiratory system silently and resulting in severe pneumonia.

Aiming to the diagnosis at initial stage; especially for asymptomatic patients, chest imaging may play a key role because the COVID-19 pneumonia is different from normal pneumonia and tuberculosis [31]. Although for lung-related disease diagnosis, chest CT imaging has been proven more effective, the CXR is preferred because comparatively it is widely available, faster and cheaper. Biomedical image segmentation and classification have become admired areas of research to make the present healthcare system more robust and responsive [2, 6]. In this context, the computing infrastructure advancements make it possible to apply the deep network approaches for complex biomedical image analysis tasks and especially, variants of

convolution neural networks (CNN) are found very effective and efficient feature learning approaches in the field of biomedical image analysis [24, 40].

Recently, several deep-learning based COVID-19 detection techniques have been proposed [31] using collection of pneumonia CXR images from different datasets [10], [18], [27], [37]. The limited availability of COVID-19 positive samples may lead to biased outcome therefore, oversampling is opted as a solution by state-of-the-art deep learning approaches. The oversampling of medical images is not always a good practice to increase the classification accuracy, hence the present paper proposes a novel one-class classification (OCC) approach for COVID-19 diagnosis in the presence of a limited number of COVID-19 positive training samples. The identification of COVID-19 infection or pneumonia in CXR images is an anomaly detection task, which is specifically an one-class classification problem. It is evident that OCSVCs (support vector data description (SVDD) and OCSVM) are more suitable for anomaly/novelty detection tasks, where the negative class samples are totally absent [3]. The OCSVCs are very sensitive to the noise, therefore the present research proposes an extended version of OCSVM [34] under the supervision of pinball loss function [17], named as PB-OCSVM. In present research work, CXR images of COVID-19 [10], radiological society of North America (RSNA) images [37] and U.S. national library of medicine (USNLM) collected montgomery country - NLM(MC) [19] datasets are utilized for experiments to justify the workability of proposed model. For experiments it is assumed that only COVID-19 infected CXR images are available for training, whereas other type of pneumonia and normal CXR images are totally absent at the time of training. The proposed PB-OCSVM possesses the following advantages:

- It is less sensitive to the noise compared to conventional OCSVM.
- Its computational complexity is nearly equal to the conventional OCSVM, because pinball loss function does not add any extra computation overhead.
- PB-OCSVM reduces to the original OCSVM when pinball loss parameter tend to '1', and thus OCSVM can be considered as a special case of the proposed PB-OCSVM.
- PB-OCSVM computes a hyperplane similar to original OCSVM.

The rest of the manuscript is organized as follows: Section 2 discusses the recent research contributions and Section 3 describes the utilized datasets. The proposed methodology is discussed in Section 4 whereas the experiments and results are discussed in Section 5. Finally, the concluding remarks and future aspects are discussed in Section 6.

2. Related work

CXR images are always preferred for disease diagnosis due to wide availability and easy handling of X-ray machines. As discussed in the preceding section, the conventional way of testing COVID-19 via RT-PCR technique is a time taking and error

prone process, especially for asymptomatic patients. Hence, the CXR image analysis is claimed to be a more robust alternate way to diagnose COVID-19. Concerning the same, several binary and multi-class classifications based deep learning models have been reported using collection of multiple CXR datasets [10], [18], [27], [37] with a common objective to develop a fully automatic diagnosis system, because the manual analysis is a time consuming process and require radiology experts. In this context, it is also evident with deep insight into literature that one-class classification approaches are not yet explored. The present research considers the diagnosis of COVID-19 infection as an anomaly or novelty detection task and proposes a novel variant of OCSVM named pin-ball OCSVM (PB-OCSVM), capable to work with limited number of training samples and robust against the noise. The proposed approach eliminates the need of oversampling of original samples as preferred by all deep learning approaches. This section gives comprehensive review of the most recently proposed deep learning approaches for COVID-19 diagnosis using CXR images.

2.1. *Deep learning approaches*

In the battle against the COVID-19 pandemic, researchers are continuously proposing artificial intelligence (AI) based quicker and alternate approaches to identify COVID-19 infection from CXR images. It is evident with deep insight into recent COVID-19 related publications that deep learning models are capable to diagnose infections using CXR images, that supplant the conventional testing methods and helpful to reduce the growing burden on radiologists during this pandemic. It is also true that there is always a chance of infection while collecting samples for swab test, whereas CXR analysis is comparatively safer and easily manageable. Seeking to identify COVID-19 infection using CXR images, Narin et al. [26] proposed deep CNN based ResNet50, InceptionV3 and Inception-ResNetV2 pre-trained transfer models. Later, Apostolopoulos et al. [4] applied state-of-the-art CNN architectures based on transfer learning approaches for CXR image classification and for experiments, two different datasets have been used where images belong to three different categories: COVID-19, viral/bacterial pneumonia and normal cases. In the race of COVID-19 diagnosis, Khalifa et al. [20] offered generative adversarial network (GAN) based approach using CXR images while ensuring the robustness against the over-fitting by generating more images. The utilized dataset contains 5863 CXR samples of two classes: pneumonia and normal. For experiments, the GoogLeNet, Squeeznet, Resnet18 and AlexNet deep learning architectures have been experimented along with GAN, where Resnet18 along with GAN outperformed other models. Further, Sethy et al. [35] introduced deep feature based support vector machine (SVM) model for COVID-19 infection classification of CXR images. It is found that SVM outperformed other models with extracted features from ResNet50. Later, effectiveness of ResNet-50 architecture was proved by Bukhari et al. [5] under three categories of CXR images: non-COVID pneumonia, normal and COVID-19

pneumonia. The experimental results show that this approach efficiently diagnose COVID-19 infection.

As an alternate way of COVID-19 infection detection, a customized ResNet-50 CNN based deep learning architecture named 'COVIDResNet' has been proposed [13]. For experiments the input images are progressively resized to $128 \times 128 \times 3, 224 \times 224 \times 3$ and $229 \times 229 \times 3$ pixels and the network is fine-tuned at each stage with automatic learning rate selection, whereas the proposed model attained higher accuracy with minimum computational complexity. Later, Zhang et al. [46] proposed a novel deep anomaly detection method for quicker and early diagnosis. To validate the performance of the proposed approach, CXR images of two different categories have been considered: COVID-19 positive cases and other pneumonia. To deal the class imbalance problem in the collected samples, authors proposed a CXR based COVID-19 screening model through anomaly detection task [30]. Most recently, Punn and Agarwal [31] proposed a transfer learning based deep learning architecture for COVID-19 diagnosis, whereas the class imbalance problem was handled via random oversampling and weighted class approaches have been used. The experimental results show that the proposed approach outperformed other state-of-the-art approaches.

Wang et al. [42] proposed a deep CNN for COVID-19 diagnosis called "COVID-Net" and experimented over COVIDx dataset. The proposed COVID-Net was pre-trained on the ImageNet dataset and then trained on the COVIDx dataset using the Adam optimizer. Performance (Multi-class classification) of this approach was compared with VGG19 and ResNet-50. Afterwards, Yujin et al. [28] proposed a patch-based ResNet-18 to work with minimum number of trainable parameters, where a novel probabilistic gradient-weighted class activation map (Grad-CAM) method was also proposed that takes into account of patch-wise disease probability in generating global saliency map. The proposed method outperformed COVID-Net when the task was considered as multi-class classification problem. Later, Afshar et al. [1] proposed a capsule network based approach for COVID-19 identification from CXR images and named it COVID-CAPS, where to deal class imbalance problem more weight is given to positive samples in the modified loss function. Experimental results show that the proposed method outperformed state-of-the-art approaches. In this context, Hall et al. [14] considered COVID-19 identification a binary classification task and proposed an ensemble of the three types of CNN classifiers for COVID-19 diagnosis. In this approach pre-trained ResNet50 and VGG-16 models work with a newly proposed small CNN in presence of balanced dataset obtained by image augmentation.

From above discussion and deep insight into literature, it is evident that almost all the existing solutions for COVID-19 diagnosis reported till date follow the deep network architectures. Every deep learning model is data hungry and consumes massive computing resources as well as time. Hence, a faster diagnosis process is needed due to the rapid growth of COVID-19 cases and limited availability of healthcare experts (especially doctors and pathologists). Meanwhile, due to limited

6   *S. K. Sonbhadra et al.*

availability of COVID-19 CXR images, class imbalance appears the most critical issue that has been addressed by all recently proposed approaches with different solutions. Therefore, a more robust solution is desired, capable of working in the presence of limited number of positive class samples whereas the other class samples are absent. Naturally, the diagnosis of COVID-19 cases can be formulated as an anomaly or novelty detection task where only COVID-19 positive samples are available for training. With this notion, following are the key contributions of the present research:

- Formulated the COVID-19 diagnosis as a one-class classification problem.
- Only COVID-19 positive CXR images are considered as training samples.
- All training and other class samples are treated as test samples.
- Performance comparison with recent deep learning approaches is done with following parameters: accuracy, precision, specificity, sensitivity and confidence_interval.
- Experiments are performed in noisy environment to show the robustness of the proposed model.
- To show generalization performance of proposed model, 10 benchmark UCI datasets are also experimented.

3. Dataset Description

In present research, CXR images of COVID-19 [10], radiological society of North America (RSNA) [37] and U.S. national library of medicine (USNLM) collected Montgomery country - NLM(MC) [19] datasets are collectively utilized for experiments to compare the performance of proposed method with existing state-of-the-art approaches. The COVID-19 dataset [10] contains CXR images of following pneumonia classes: SARSr-CoV-1 (SARS), SARSr-CoV-2 (COVID-19), Pneumocystis spp., Streptococcus spp. and ARDS from multiple public domain resources available without infringing patient's confidentiality (Fig. 1 (b)). The dataset includes the statistics up to August 2020 with metadata: patient ID, offset, sex, age, finding, survival, view, modality, etc. For experiments other samples from COVID-19 infected images are considered as another class named as 'other pneumonia'. The current version of the dataset is more rich in number of samples and classes, but to show the effectiveness of the proposed approach, the older version is used.

Another dataset used in this research has been introduced under RSNA pneumonia diagnose competition, which is a collection of 30,000 images chosen from the NIH CXR14 dataset [37]. Among these samples, 15,000 images are pneumonia infected and out of remaining 15,000 samples, 7,500 had no findings whereas other 7,500 had different symptoms from pneumonia. Medical experts and radiologists annotated these images. Fig. 1(c) shows sample instances. Initially, this dataset was published with 25,684 training and 1,000 test images and later, the 1000 test samples were merged to the existing training set to form 26,684 training images whereas 3,000 new images were introduced as test set. For more robust perfor-
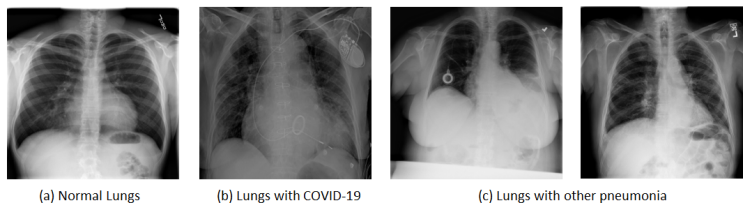
(a) Normal Lungs   (b) Lungs with COVID-19   (c) Lungs with other pneumonia

Fig. 1. Sample CXR images.

Table 1. Summary of utilized datasets.

Dataset	Behaviour	CXR images	Total samples
COVID-19	COVID-19	108	153
	Other pneumonia	45	
RSNA	Normal	453	923
	Other pneumonia	457	
NLM(MC)	Tuberculosis	58	138
	Normal	80	

mance evaluation and comparative analysis, NLM(MC) [19] dataset is also utilized that consists of 138 CXR images of classes: tuberculosis and normal. Table 1 shows the description of fused dataset used in present research. The dataset is composed of 1214 posteroanterior CXR images with classes: COVID-19 (108), other pneumonia (515), tuberculosis (58) and normal (533).

4. Proposed approach

From preceding sections, it is inferred that deep learning approaches are playing a decisive role in advanced biomedical image analysis tasks. Several deep learning models have been proposed for COVID-19 diagnosis but suffer with following key issues:

- These models need massive computation power for training.
- Fails to work if samples of only one class (target class) are available for training. This phenomenon is the motivation of this research.
- May give biased outcome if the dataset suffers with class imbalance problem.

The above issues are identified as the research gap and the present article proposes a one-class classification approach to solve these issues. The conventional OCSVM (a variant of OCSVCs) and the proposed PB-OCSVM are used as a one-class classifiers, capable of working with only target class (also known as class of interest (CoI)) samples. The OCC problem is different from conventional binary and multi-class classification tasks, where only target class is well defined and the
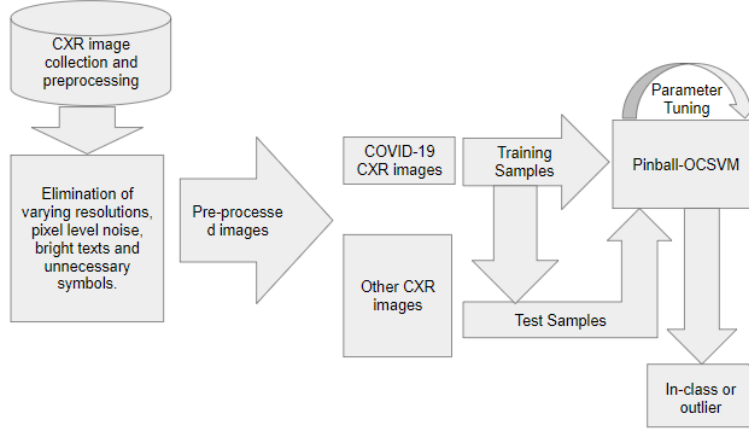
8   *S. K. Sonbhadra et al.*

Fig. 2. Schematic of proposed model

other class samples are either totally absent or poorly sampled. In present research, only COVID-19 infected CXR images are used as training samples and the performance evaluation is done in two phases: considering all COVID-19 samples as test samples and considering other class images as test samples. The aim of the present research is to ensure maximum learning ability in presence of limited number of CXR images with minimum false-rejection and false-acceptance. Fig. 2 schematic of the proposed approach where it is assumed that only the target class samples are available for training. Initially, data pre-processing is performed to remove the noise from the images and then the processed images are fed to PB-OCSVM for training. After training, test samples (in-class and other class samples) are tested.

4.1. *Data pre-processing*

In this article, COVID-19 infected posteroanterior CXR images of COVID-19 dataset [10] are used to train the PB-OCSVM. For performance evaluation of the proposed model, the other pneumonia samples from COVID-19 dataset, RSNA [37] and NLM(MC) [19] datasets are used as test samples along with COVID-19 samples. The RSNA and NLM(MC) datasets contain samples of pneumonia and tuberculosis respectively along with normal cases. It is evident that the CXR images in the aggregated dataset suffers with undesired artifacts such as varying resolutions, pixel level noise, bright texts and unnecessary symbols, hence pre-processing is a necessary step for further analysis. To overcome from the textual and symbolic noise, in-painting is performed with the image mask generated using binary thresholding [29] (Eq. 1), whereas image resizing is performed for a fixed size resolution of $331 \times 331 \times 3$ (where 3 is number of channels) to work upon state-of-the-art models.

$$N(x,y) = \begin{cases} Th_{max}, & \text{i(x,y)} \geq Th_{min} \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where for input image $i(x,y)$, the mask is designed with minimum and maximum threshold values defined as Th_{min} and Th_{max} respectively. Even after filtering out undesired information, there is always a possibility of uncertainty at the deep pixel level representation [15].

To preserve the original pixel value distribution, following two techniques are experimented to generate final pre-processed images: the adaptive total variation method [39] and automatic color equalization (ACE) approach [33]. Images of size $331 \times 331 \times 3$ are used by these methods to generate final images.

4.1.1. *Adaptive total variation method*

Let, I is a grayscale image and γ is a bounded set over $\mathbb{R}^2$, such that $\gamma \subset \mathbb{R}^2$, denoising image $\mathcal{D}$ that closely matches to observed image $x = (x_1, x_2) \, \epsilon \, \gamma$ - pixels, given as:

$$\mathcal{D} = \arg\min_{\mathcal{D}} \left(\int_\gamma (\mathcal{D} - I.\ln\mathcal{D})dx + \int_\gamma (\omega(x)|\Delta\mathcal{D}|dx) \right) \tag{2}$$

where $\omega(x) = \frac{1}{1+p \mod G_\sigma * \Delta \mathcal{D}'}$, G_σ is the Gaussian kernel, σ is variance , $p > 0$ is contrast parameter and $*$ is convolution operator.

4.1.2. *Automatic color equalization*

This method is based on a computational model of the human visual system that merges the two basic global equalization mechanisms: "Gray World" (GW) and "White Patch" (WP). ACE is able to adapt widely varying lighting conditions and to extract visual information from the environment efficaciously. This is a two stage process where initially, the GW and WP approaches are merged via a type of lateral inhibition mechanism, weighted by pixel distance that results in local-global filtering. Whereas the second stage maximizes the image dynamic via normalizing the white at a global level.

Let the input image is I with c channels and initially, the chromatic/spatial adjustment produces as output image R in first stage, where every pixel is recomputed according to the image content. Each pixel p of the output image R is computed separately for each channel c as follows:

$$R_c(p) = \sum_{j \neq p} \frac{r(I_c(p) - I_c(j))}{d(p,j)} \tag{3}$$

10   *S. K. Sonbhadra et al.*

where $I_c(p) - I_c(j)$ decides the lateral inhibition mechanism, the distance $d(.)$ weights the amount of local or global contribution and $r(.)$ is the relative lightness appearance of the pixel. Like human visual system, the distance $d(.)$ weights the global and local filtering effect. Euclidean distance is considered for distance calculation. Relative pixel influence of each pixel is controlled by $r(.)$ and $d(.)$ for the spatial channel lightness adjustment. The variation of the slope of the function $r(.)$ acts as a contrast tuner and in the present research Signum function is used. Finally, the second stage linearly scales the range of values in R_c independently in the relative channel c into the range $[0; 255]$ using the formula:

$$O_c(p) = round[s_c R_c(p) - m_c] \tag{4}$$

for each pixel p where s_c is the slope of the segment $[(m_c, 0); (M_c, 255)]$, with $M_c = max_p R_c(p)$ and $m_c = min_p R_c(p)$. To make this process more robust, dynamic of the final image is always centered around the medium gray via means of White Patch/Gray World (WP/GW) scaling as follows:

$$O_c(p) = round[127.5 + s_c R_c(p)] \tag{5}$$

For experiments, Euclidean distance and WP/GW scaling are used and s_c is set to 20 for saturation function.

The complete data pre-processing operation is illustrated in Fig. 3, where the histogram of resulting distributed pixels during every step of pre-processing is also
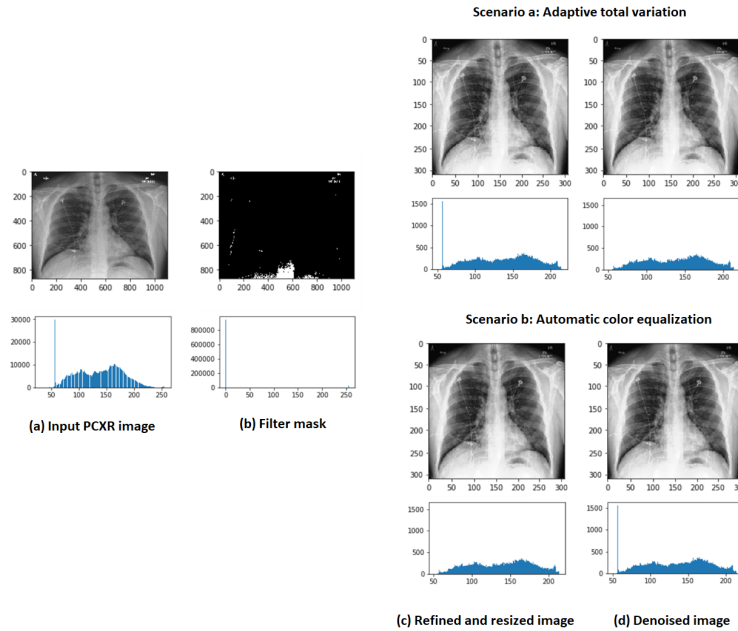

Fig. 3. Data pre-processing phases of chest X-ray image.

shown. It is evident that the opted pre-processing approach is capable to eliminate the irregular intensities while preserving the original nature of the pixel distribution. After pre-processing of CXR images, intensive experiments are performed with PB-OCSVM, OCSVM and state-of-the-art deep network models. For PB-OCSVM and OCSVM, the COVID-19 positive samples are considered as training samples whereas deep learning models are trained with multi-class samples (80% from each class) obtained via respective class balancing techniques. For OCC approaches, all in-class and other class samples are treated as test samples and for deep learning models all training, remaining in-class and other class samples are treated as test samples.

4.2. *Preliminaries*

This section gives brief description of OCSVM, variants of OCSVM with different loss functions and Pinball-SVM. This discussion will help to understand the formulation of proposed PB-OCSVM.

4.2.1. *OCSVM*

OCC algorithms are specifically used for anomaly/novelty detection tasks; especially when the non-target class is either ill-defined or totally absent. For unbiased operation of binary or multi-class classifiers, presence of two well-defined classes is necessary; but if the test sample belongs to unknown class, the classifier may exhibit biased behaviour. In such scenarios, OCC techniques are proven as robust solutions, majorly applicable for concept learning and outliers/novelty detection [21].

Tax et al. [41] proposed a novel OCC model termed as SVDD, where the target class samples are enclosed by a hypersphere, and the boundary points are called support vectors. If a data sample falls outside of the hypersphere, SVDD treats it as outlier and rejects (Fig. 4). The objective function of SVDD is defined as:

$$L(R, a, \alpha_i, \gamma_i, \xi_i) = R^2 + C \sum_i \xi_i - \sum_i \alpha_i \{R^2 + \xi_i - (\| x_i \|^2 - 2a.x_i + \| a \|^2)\} - \sum_i \gamma_i \xi_i \tag{6}$$

subject to: $\| x_i - a \|^2 \le R^2 + \xi_i, where \ \xi_i \ge 0 \ \forall$ i
where radius of the hypersphere is R, x_i and a are outlier and center of hypersphere respectively, the parameter C controls the trade-off between the erros and volume, and ξ is slack variable to penalizes the outliers. With the Lagrange multipliers $\alpha_i \ge 0$, $\gamma_i \ge 0$ where $i \in \{1, 2, \ldots, N\}$, the objective is to minimize R of the hypersphere while covering all target class samples with some penalty for outliers. By putting partial derivatives to zero and substituting the constraints into Eq. 6, following is obtained:
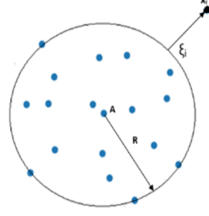
12   *S. K. Sonbhadra et al.*

Fig. 4. Support vector data description (SVDD).

$$L = \sum_i \alpha_i(x_i, x_i) - \sum_{i,j} \alpha_i \alpha_j(x_i, x_j) \tag{7}$$

If the description value of a test sample x_i is greater than C then it is treated as an outlier. Kernels can be used to reformulate the SVDD as follows:

$$\|\phi(x) - a\|^2 \leq R^2 \tag{8}$$

The SVDD outcome can be computed as follows:

$$R^2 - \|\phi(x) - a\|^2 \tag{9}$$

The output of Eq. 9 is positive for samples inside the boundary and negative for an outliers.

Later, Schlökopf et al. [34] proposed one-class support vector machine (OCSVM) as an alternate OCC approach to SVDD, where a hyperplane separates the target class samples from outliers as shown in Fig. 5. In OCSVM, class of interest (CoI) samples are separated by a hyperplane with the maximal margin from the origin, whereas the negative class samples reside in the subspace of the origin. The OCSVM is defined as following quadratic equation (Eq. 10):

$$\max_{w,\xi,\rho} \frac{1}{2} \| w \|^2 + \frac{1}{vN} \sum_i^N \xi_i - \rho \tag{10}$$

subject to: $w.\phi(x_i) \geq \rho - \xi_i$ and $\xi_i \geq 0$ $\forall\, i \in \{1, 2, \ldots n\}$.
where in feature space the data sample x_i is represented by ϕ and the outlier is penalized by slack variable ξ_i. $v \, \epsilon \, (0,1]$ decides the lower bound on the number of support vectors and upper bound on the fraction of outliers.

The dual optimization problem of Eq. 10 is defined as follows:

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j K(x_i, x_j) \tag{11}$$
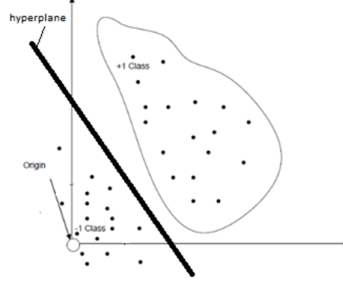
Fig. 5. Working of OCSVM.

subject to: $0 \leq \alpha_i \leq \frac{1}{vN}$, $\sum_{i=1}^{N} \alpha_i = 1$, $i = \{1, 2, \ldots, n\}$.

where $\alpha = [\alpha_1, \alpha_2, \ldots \alpha_N]^T$ and α_i is the Lagrange multiplier, whereas the weight-vector w can be expressed as:

$$w = \sum_{i=0}^{N} \alpha_i \phi x(i) \tag{12}$$

ρ is the margin parameter and computed by any x_i whose corresponding Lagrange multiplier satisfies $0 < \alpha_i < \frac{1}{vN}$

$$\rho = \sum_{j=1}^{N} \alpha_j K(x_j, x_i) \tag{13}$$

The decision function with kernel expansion can be written as:

$$f(x) = \sum_{i=1}^{N} \alpha_i K(x_i, x) - \rho \tag{14}$$

Label of any test sample x can be decided with the following:

$$\hat{y} = sign(f(x)) \tag{15}$$

where $sign(.)$ is sign function.

It is observed that both SVDD and OCSVM perform same with Gaussian kernel and origin plays a key role because all outliers reside there. Norm of the centre of SVDD is equal to margin of a hyperplane of OCSVM in unit norm feature space is [22] as depicted in Fig. 6(a). Thus, reformulation of SVDD can be done by a hyperplane as follows:

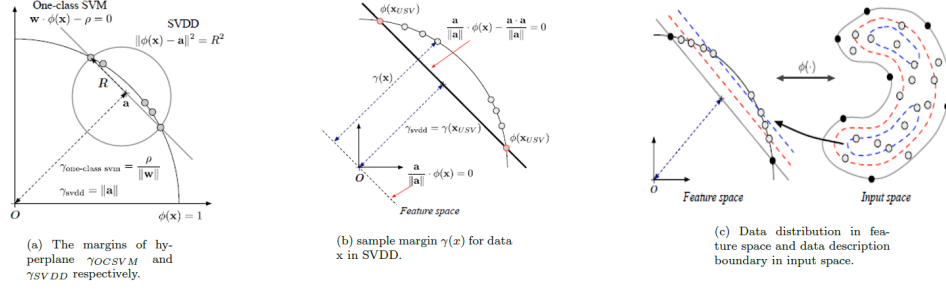$$\| \phi(x) - a \|^2 \leq R^2 \Leftrightarrow w_{SVDD}.\phi(x) - \rho_{SVDD} \geq 0 \tag{16}$$

14   *S. K. Sonbhadra et al.*

(a) The margins of hyperplane γ_{OCSVM} and γ_{SVDD} respectively.

(b) sample margin $\gamma(x)$ for data x in SVDD.

(c) Data distribution in feature space and data description boundary in input space.

Fig. 6. Representation of SVDD and OCSVM in unit norm space [35].

where a represents the the centre the hypersphere. w_{svdd} and ρ_{svdd} are the normal vector and the bias respectively of SVDD hyperplane and defined as below:

$$w_{SVDD} = \frac{a}{\parallel a \parallel}, \qquad \rho_{SVDD} = \parallel a \parallel \tag{17}$$

In feature space, the virtual hyperplane passes through the origin and the sample margin is defined by its distance from the image of the data x as shown in Fig. 6(b). The SVDD's sample margin is defined as below:

$$\gamma_{SVDD}(x) = \frac{a.\phi(x)}{\parallel a \parallel} \tag{18}$$

where $\gamma(x)$ is the image of data x in feature space and the sample margin of OCSVM is defined as follows:

$$\gamma_{OCSVM}(x) = \frac{w.\phi(x)}{\parallel w \parallel} \tag{19}$$

Because data samples exist on the surface of a unit hypersphere, the maximum and minimum value of sample margin is defined as '1' and '0' respectively as follows:

$$0 \leq \gamma(x) \leq 1 \tag{20}$$

Also, sample margin of unbounded support vectors x_{USV} ($0 < \alpha_{x_{USV}} < \frac{1}{\upsilon N}$) are the same as the margin of hyperplane, therefore:

$$\gamma(x_{USV}) = \gamma_{SVDD} = \parallel a \parallel \tag{21}$$

$$\gamma(x_{USV}) = \gamma_{OCSVM} = \frac{\rho}{\parallel w \parallel} \tag{22}$$

In feature space, the distribution of samples is represented by the sample margiin. Fig. 6(c) shows the distribution of sample margin of training data and hyperplane of OCSVM in feature space. It is also observed that for RBF kernel both SVDD and OCSVM perform equally, hence in the present research, OCSVM is chosen for enhancement with the help of pinball loss function.

4.2.2. *OCSVM with other loss functions*

For every machine learning tasks such as classification, regression, etc., the loss functions plays vital role. Consider the training set $S = \{x_1, x_2, \dots x_n\}$, where $x_i \in R^n$ are inputs and target class y is defined as $y \in \{1\}$. Let, $f : R^n \to R$ is a mapping from $x_i \in R^n$ to $y \in \{1\}$. During training and testing the predicted output value of samples is supervised by the associated loss functions.

- *OCSVM with hinge loss function:*
  Like other machine learning tasks the hinge loss for OCSVM [45] is defined as follows:

$$H_{loss}(z_i) = max(0, \rho - z_i) \qquad (23)$$

  This unbounded loss function is capable to deal with the shortest distance of the target class samples, but sensitive to the noise. The hinge loss OCSVM can be written as:

$$\max_{w, \xi, \rho} \frac{1}{2} \parallel w \parallel^2 - \rho + \frac{1}{\upsilon N} \sum_{i}^{N} H_{loss}(z_i) \qquad (24)$$

  where $z_i = w\phi(x_i)$. The constraints defined in Eq. 24 are integrated into $H_{loss}(z_i)$. The samples satisfying $z_i = w.\phi(x_i) \geq \rho$, are positioned above the hyperplane and no penalty is added, hence $H_{loss}(z_i) = 0$, whereas samples satisfying $z_i = w\phi(x_i) < \rho$ are located on the other side of the hyperplane and some penalty should be inflicted, therefore $H_{loss}(z_i) = \rho - z_i > 0$, which says that if the samples move far away from the hyperplane the penalty increases.
- *OCSVM with ramp loss function:*
  As described above, the hinge loss function is unbounded and sensitive to the noise, the ramp loss function was introduced to OCSVM to make the model more robust [45]. The ramp loss function is described as follows:

$$R_{loss(\rho,l)}(z_i) = \begin{cases} 0, & z_i \geq \rho \\ \rho - z_i, & \rho l < z < \rho \\ \rho - l\rho, & z_i \leq \rho l \end{cases} \qquad (25)$$

16   *S. K. Sonbhadra et al.*

where $0 < l < 1$. When $z_i > l\rho$, the ramp loss is same to the hinge loss; whereas for $z_i \le l\rho$, the ramp loss is a constant that is an advantage over hinge loss which increases as z_i decreases. This loss function is capable to deal with the shortest distance of the target class samples and less sensitive to the noise. The ramp loss OCSVM can written as:

$$\max_{w,\xi,\rho} \frac{1}{2} \parallel w \parallel^2 -\rho + \frac{1}{\upsilon N} \sum_{i}^{N} R_{loss}(z_i) \tag{26}$$

where $z_i = w\phi(x_i)$. Due to non-convexity of ramp loss the OCSVM problem is no longer a convex optimization. This problem is solved by the phenomenon that the R_{loss} is actually the difference between two hinge loss functions and the modified formulation of the ramp loss OCSVM consists of the convex part and the concave part. This new problem is solved by "Concave-Convex Procedure" (CCP) procedure [11].

• *Pin ball SVM:* The pinball loss function is given as follows, which can be regarded as a generalized l_1 loss.

$$P_\tau(u) = \begin{cases} u, & u \ge 0 \\ -\tau u, & u < 0 \end{cases} \tag{27}$$

where range of τ is [0, 1] [17]. For quantile regression [23], [9], the pinball loss P_τ has been successfully applied.

Huang et al. [17] derived quantile classification using the idea of support vector machine (SVM). It has been shown that pinball and hinge loss SVM have similar consistency property and computational complexity, moreover the pinball loss SVM is less sensitive to noise.

Huang et al. [17] introduced pinball loss into the SVM. Consider the data points to be classified denoted by the set T. To solve the classification task, $w \in R^n$ and $b \in R$ must be identified such that:

$$\begin{aligned} w^T\phi(x_i) + b \ge 1 & \quad \text{for } y_i = 1 \\ w^T\phi(x_i) + b \le -1 & \quad \text{for } y_i = -1 \end{aligned} \tag{28}$$

where $\phi(.)$ is a nonlinear mapping from the input space to a new feature space. The optimal hyperplane $w^T\phi(x) + b = 0$, lies exactly between the supporting parallel hyperplanes given by:

$$w^T\phi(x) + b = 1 \text{ and } w^T\phi(x) + b = -1 \tag{29}$$

and separates the binary class samples from each other with a margin of $\frac{1}{\|w\|}$ on each side. Data points residing on the supporting hyperplanes (Eq. 29) are termed as support vectors. The classifier is obtained by maximizing the margin. If the data points of two different classes are not linearly separable

in feature space then to separate all the data points correctly, pinball loss $P_\tau(x, y, f(x))$ is introduced to allow the existence of data points that violate the constraints $y_i(w^T \phi(x_i) + b) \geq 1$. Finally, the nonlinear Pin-SVM is formulated as follows:

$$\max_{w,b} \frac{1}{2} \parallel w \parallel^2 + c \sum_{i=0}^{t} P_\tau(z_i) \qquad (30)$$

After employing the pinball loss function in Eq. 30, following quadratic programming problem (QPP) come into existence for Pin-SVM:

$$\max_{w,\xi,\rho} \frac{1}{2} \parallel w \parallel^2 + c \sum_{i=1}^{t} \xi_i \qquad (31)$$

s.t.

$$
\begin{aligned}
y_i(w^T \phi(x_i) + b) &\geq 1 - \xi_i, \\
y_i(w^T \phi(x_i) + b) &\leq 1 + \frac{\xi_i}{\tau}, i = 1, 2, \ldots N
\end{aligned}
\qquad (32)
$$

where $\xi = (\xi_1, \xi_2, ..., \xi_N)^T$ is a slack variable and $c > 0$ is a penalty parameter. The parameter c determines the weight between the two terms $\parallel w \parallel^2$ and $\sum_{i=1}^{t} \xi_i$. Pin-SVM has an advantage of noise insensitivity. Meanwhile, Pin-SVM has similar time complexity to that of standard SVM. Note that the second constraint of Eq. 32 becomes $\xi \geq 0$ when $\tau = 0$, and thus Pin-SVM reduces to the hinge loss SVM.

4.3. *OCSVM with pinball loss function (PB-OCSVM)*

From above discussion it is clear that the hinge loss is sensitive to the noise. To overcome from this problem and to improve the performance of OCSVM, pinball loss function [17] is fused to standard OCSVM to obtain pinball OCSVM (PB-OCSVM). This enhanced version deals with quantile distance [23] that reduces the sensitivity to the noise. After introducing pinball loss function, the conventional OCSVM can be rewritten as follows.

$$\max_{w,\xi,\rho} \frac{1}{2} \parallel w \parallel^2 - \rho + \frac{1}{\upsilon N} \sum_{i}^{N} P_\tau(z_i) \qquad (33)$$

where P_τ is pinball loss as shown in Eq. 27 with $z_i = w^T \phi(x_i)$ which helps to the slope on the target class. Substituting the pinball loss into Eq. 33, following QPP is obtained:

18   *S. K. Sonbhadra et al.*

$$\max_{w,\xi,\rho} \frac{1}{2} \parallel w \parallel^2 -\rho + \frac{1}{\upsilon N} \sum_i^N \xi_i$$

$$\text{subject to:} \qquad w.\phi(x_i) \geq \rho - \xi_i \qquad\qquad (34)$$

$$w.\phi(x_i) < \rho + \frac{\xi_i}{\tau}$$

4.3.1. *Dual problem and kernel formulation*

Now after introducing a kernel based formulation to the PB-OCSVM, the Lagrangian with $\alpha_i \geq 0$, $\beta_i \geq 0$ of Eq. 34 is as follows:

$$\mathcal{L}(w, b, \xi, \alpha, \beta) = \frac{1}{2} \parallel w \parallel^2 + c \sum_{i=1}^N \xi_i - \sum_{i=1}^N \alpha_i(w.\phi(x_i)-$$

$$\rho + \xi_i) - \sum_{i=1}^N \beta_i(w.\phi(x_i) - \rho - \frac{\xi_i}{\tau}) \qquad (35)$$

According to:
$\frac{\delta \mathcal{L}}{\delta w} = w + \sum_{i=1}^N \alpha_i \phi(x_i) = 0,$
$\frac{\delta \mathcal{L}}{\delta \rho} = \sum_{i=1}^N \alpha_i - \beta_i = 0,$
$\frac{\delta \mathcal{L}}{\delta \xi} = c - \alpha_i - \frac{1}{\tau}\beta_i = 0, \forall i = 1, 2, \ldots, N$
  the dual problem of Eq. 35 is obtained as follows:

$$\max_{\alpha,\beta} -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - \beta_i)\phi(x_i)^T \phi(x_j)(\alpha_j - \beta_j) + \sum_{i=1}^N (\alpha_i - \beta_i)$$

subject to:

$$\sum_{i=1}^N (\alpha_i - \beta_i) = 0, \qquad\qquad (36)$$

$$\alpha_i + \frac{1}{\tau}\beta_i = c, i = 1, 2, \ldots N,$$

$$\alpha_i \geq 0, \beta_i \geq 0, i = 1, 2, \ldots N$$

Introducing the positive definite kernel $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ and variables $\lambda_i = \alpha_i - \beta_i$, we get

$$\max_{\alpha,\beta} -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i K(x_i,x_j)\lambda_j + \sum_{i=1}^{N}\lambda_i$$

subject to:

$$\sum_{i=1}^{N}\lambda_i = 0, \tag{37}$$

$$\lambda_i + (1+\frac{1}{\tau})\beta_i = c, i = 1,2,\ldots N,$$

$$\lambda_i + \beta_i \geq 0, \beta_i \geq 0, i = 1,2,\ldots N$$

With kernel expansion the decision function can be defined as follows:

$$\{(x) = \sum_{i=1}^{N}\lambda_i K(x_i,x) - \rho \tag{38}$$

Finally, the test instance x can be labelled as follows:

$$\hat{y} = sign(\{(x)) \tag{39}$$

where $sign(.)$ is sign function.

It is well known that the computational complexity of OCSVM is $\mathcal{O}(N_{T_c}^3)$ where N_{T_c} is the number of the only target class samples, whereas the computational complexity of PB-OCSVM is found same as OCSVM i.e. $\mathcal{O}(N_{T_c}{}^3)$. Therefore, the complexity of the PB-OCSVM is equal to conventional OCSVM, hence pinball loss function does not increase the computation time of PB-OCSVM.

5. Experiments and results

COVID-19 positive samples of COVID-19 dataset [10] are used as training samples for OCSVM and the proposed PB-OCSVM whereas other samples of this dataset along with samples of datasets RSNA [37] and U.S. national library of medicine (USNLM) collected Montgomery country - NLM(MC) [19] are used for testing. All the images of fused dataset is pre-processed as discussed in Subsection 4.1 that outputs the images of size $331 \times 331 \times 3$, where 3 is the number of channels. For PB-OCSVM and OCSVM, the COVID-19 positive samples are considered as training samples whereas deep learning models are trained with multi-class samples (80% from each class) obtained via respective class balancing techniques. For OCC approaches, all in-class and other class samples are treated as test samples and for deep learning models all training, remaining in-class and other class samples are treated as test samples. Once training is done with suitable parameters, OCSVM and PB-OCSVM are used for testing in two scenarios: test with training samples and other samples of all three datasets. For performance evaluation five benchmark parameters have been used to show effectiveness of the proposed model: accuracy,

precision, sensitivity, specificity and confidential interval (CI) with 95% and 98% confidence (Eq. 40). Comparison of performance of OCSVM and PB-OCSVM is done with recently proposed deep learning models as shown in Table 2. It is known that all deep learning models need huge amounts of data and at the same time all classes must be well defined. Hence, it is necessary to oversampling of COVID-19 samples to any deep learning model. It is observed that all recent deep learning architectures [28], [1], [42], [4], [14] used oversampling to deal with the data imbalance problem and after massive computational efforts were able to give significant performance.

$$
\begin{aligned}
Accuracy &= \frac{TN + TP}{TN + FN + TP + FP} \\
Precision &= \frac{TP}{TP + FP} \\
Specificity &= \frac{TN}{TN + FP} \\
Sensitivity &= \frac{TP}{TP + FN} \\
Confidence\_Interval &= z * (\sqrt{accuracy * (1 - accuracy)}/N
\end{aligned} \tag{40}
$$

where z is standard deviation (1.96 for 95% confidence and 2.33 for 98% confidence) and $N = 1214$ is number of samples.

There are total four classes in fused dataset: COVID-19, normal, Tuberculosis and other pneumonia. Some existing models have been proposed considering all four classes whereas some articles considered Tuberculosis and other pneumonia classes as a single class that gives three classes: COVID-19, normal and other pneumonia. Few papers considered normal, Tuberculosis and other pneumonia as non COVID class that gives two classes COVID and non COVID. For existing deep learning architectures, experiments have been performed in three categories with consideration of 4, 3 and 2 classes. For exhaustive comparative analysis, all scenarios have also been considered for OCSVM and PB-OCSVM and it is observed that both the models exhibited the consistent performance due of their working principle. Whereas it is identified that deep learning models perform better in case of binary class classification approach. It is also observed that both adaptive total variation and ACE performed nearly equal.

5.1. *Workability of proposed PB-OCSVM in presence of noise*

To validate the robustness of the proposed PB-OCSVM against the noise, and to demonstrate its generalized behaviour experiments are performed considering following scenarios:

   a. Images obtained after inpainting operation as discussed in Subsection 4.1.

Table 2. Performance comparison.

Authors	Technique	Classes	Adaptive total variation					Automatic color equalization				
			A	P	S	Sp	95% & 98% CI	A	P	S	Sp	95% & 98% CI
Yujin et al. [28]	Fine Tuning		0.88	0.83	0.86	0.96	± (1.83% - 2.17%)	0.86	0.86	0.88	0.92	± (1.95% - 2.32%)
Afshar et al. [1]	Fine Tuning	4	0.95	0.92	0.90	0.95	± (1.10% - 1.31%)	0.94	0.94	0.92	0.91	± (1.36% - 1.58%)
OCSVM			0.97	0.97	0.93	0.96	± (0.96% - 1.14%)	0.96	0.96	0.94	0.95	± (1.10% - 1.31%)
PB-OCSVM			**0.98**	**0.99**	**0.98**	**0.98**	± (0.79% - 0.94%)	**0.97**	**0.98**	**0.98**	**0.98**	± (0.95% - 1.14%)
Wang et al. [42]	Fine Tuning		0.92	0.91	0.88	0.89	± (1.53% - 1.81%)	0.92	0.91	0.9	0.91	± (1.53% - 1.81%)
Apostolopoulos et al. [4]	Fine Tuning	3	0.87	0.93	0.92	**0.98**	± (1.89% - 2.25%)	0.89	0.94	0.93	0.95	± (1.76% - 2.09%)
OCSVM			0.97	0.97	0.93	0.96	± (0.96% - 1.14%)	**0.96**	**0.96**	0.94	0.96	± (1.10% - 1.31%)
PB-OCSVM			**0.98**	**0.99**	**0.98**	**0.98**	± (0.79% - 0.94%)	**0.96**	**0.96**	**0.95**	**0.97**	± (1.10% - 1.31%)
Hall et al. [14]	Fine Tuning		0.91	0.88	0.88	0.93	± (1.61% - 1.91%)	0.89	0.89	0.88	0.93	± (1.76% - 2.09%)
Apostolopoulos et al. [4]	Fine Tuning	2	**0.98**	0.97	0.92	**0.98**	± (0.79% - 0.94%)	0.96	0.95	0.92	0.95	± (1.10% - 1.31%)
OCSVM			0.97	0.97	0.93	0.96	± (0.96% - 1.14%)	0.95	0.95	0.94	0.94	± (1.22% - 1.46%)
PB-OCSVM			**0.98**	**0.99**	**0.98**	**0.98**	± (0.79% - 0.94%)	**0.97**	**0.98**	**0.98**	**0.98**	± (0.95% - 1.14%)

* A- Accuracy, P- Precision, S- Sensitivity, Sp- Specificity

    b. Images generated after introducing Gaussian, laplacian and uniform noise to the final images obtained by adaptive total variation method.

    c. Experiments with 10 benchmark UCI datasets.

For scenarios *a* and *b* all state-of-the-art methods are evaluated along with proposed method whereas for scenario *c* conventional OCSVM and proposed PB-OCSVM are evaluated. In scenario *a* the images obtained after binary thresholding as discussed in Subsection 4.1 are used for computation. The partially processed images are used by all the above discussed classifiers. Whereas, to illustrate the effectiveness of the proposed method, the output images obtained after pre-processing are further fused with Gaussian, laplacian and uniform noise are used in scenario *b*. Table 3 shows the comparative performance of state-of-the-art approaches and proposed approach (best results are represented in bold). To tune the hyperparameters of OCSVM and PB-OCSVM, grid search mechanism is used. Results show that the in the presence of noise, the proposed method outperformed compared to deep learning models and conventional OCSVM.

Table 3. Performance comparison in noisy environment

Authors	Classes	Scenario-a (Partially processed images)				Scenario-b (in presence of noise) AUC		
		A	P	S	Sp	Gaussian	Laplacian	Uniform
Yujin et al. [28]		0.82	**0.86**	0.87	**0.89**	0.83	0.84	0.86
Afshar et al. [1]	4	0.83	0.81	0.86	0.82	0.79	0.78	0.81
OCSVM		0.82	0.81	0.85	0.83	0.84	0.83	**0.88**
PB-OCSVM		**0.85**	**0.86**	**0.88**	**0.89**	**0.88**	**0.89**	**0.88**
Wang et al. [42]		0.82	0.82	0.85	0.84	0.84	0.85	0.84
Apostolopoulos et al. [4]	3	0.85	0.84	0.85	0.81	0.75	0.79	0.81
OCSVM		0.82	0.81	0.85	0.83	0.83	**0.89**	0.84
PB-OCSVM		**0.88**	**0.86**	**0.89**	**0.91**	**0.89**	**0.89**	**0.88**
Hall et al. [14]		0.81	0.83	0.82	0.84	0.81	0.82	0.84
Apostolopoulos et al. [4]	2	0.79	0.82	0.84	0.85	0.76	0.78	0.81
OCSVM		0.86	0.86	0.85	0.87	0.83	0.86	0.84
PB-OCSVM		**0.91**	**0.89**	**0.91**	**0.92**	**0.88**	**0.89**	**0.89**

* A- Accuracy, P- Precision, S- Sensitivity, Sp- Specificity

To illustrate the generalized effectiveness of the proposed model, 10 benchmark UCI datasets are experimented with OCSVM and PB-OCSVM. Table 4 shows the description of datasets used and AUC score of both the classifiers. Experimental results show that the PB-OCSVM performs marginally better than the conventional OCSVM. To ensure the workability in minimum number of samples, 80% of the target class samples are randomly chosen for training whereas all target class samples and other class instances are selected for testing as done in earlier experiments.

Table 4. Experimental results with UCI datasets

Dataset	N_{Target}	$N_{Outliers}$	Attributes	$N_{Training}$	N_{Test}	AUC	
						OCSVM	PB-OCSVM
Blood Transfusion	178	570	4	142	606	**0.86**	0.85
Wholesale Customers	298	142	7	238	202	0.94	**0.96**
Breast Cancer	77	186	9	62	201	0.93	**0.96**
Glass	70	77	10	56	91	**0.98**	**0.98**
Heart	120	150	13	96	174	0.95	**0.96**
Climate Model	294	46	18	235	105	0.94	**0.98**
Hepatitis	123	32	19	98	57	**0.95**	**0.95**
Parkinsons	147	48	22	118	77	0.97	**0.98**
QSAR biodegradation	356	699	41	285	770	**0.95**	0.94
Sonar	111	97	60	89	119	0.93	**0.95**

* N_{Target}: number of target samples, $N_{Outliers}$: number of other class samples etc.

6. Conclusion

The present research proposes a novel pinball OCSVM (PB-OCSVM) for early-detection of COVID-19 in presence of limited number of samples. Recently, several deep learning based automated approaches have been proposed for diagnosis of COVID-19 using CXR images. All these solutions need enormous multi-class data samples for unbiased operation. Due to limited availability of CXR images, it is evident that all deep learning approaches use collection of multiple datasets and prefer oversampling strategies to ensure smooth functioning. The COVID-19 is a newly known deadly pandemic and its precise characteristics are still unknown; therefore, oversampling must not be considered as a concrete solution to achieve higher diagnosis accuracy.

In present research, the COVID-19 diagnosis is considered as an anomaly/ novelty detection task and a novel variant of OCSVM is proposed under the supervision of pinball loss function called PB-OCSVM capable to work with limited number of COVID-19 infected CRX images. The proposed model does not add extra computation overhead and ensures quicker diagnosis. For experiments collection of multiple datasets are used where the available COVID-19 infected X-ray images are used for training whereas all in-class and other class samples are treated as test samples. The same fused dataset is used to evaluate state-of-the-art deep learning models and conventional OCSVM. Experimental results show that PB-OCSVM outperformed state-of-the-art deep learning models and conventional OCSVM. Meanwhile, it is

also evident that the proposed model ensures negligible false-positive and false-negative rates compared to OCSVM and attains significantly better accuracy, precision, specificity and sensitivity.

To validate the performance and robustness of PB-OCSVM, experiments are performed with noisy images. Two noisy scenarios are considered for experiments and results show that PB-OCSVM performs better in both the scenarios compared to state-of-the-art approaches. For generalized performance evaluation of proposed model, experiments are also performed with 10 benchmark UCI datasets. Experimental results proves that PB-OCSVM performs significantly better than conventional OCSVM.

The disease diagnosis in healthcare domain is especially an anomaly or novelty detection task where one-class classification approaches are proven more promising compared to conventional binary or multi-class classification approaches. In this context, it is believed that the application of proposed model is not limited and can also be extended to other disease diagnosis tasks. Meanwhile, the same work can also be extended to other application domains for anomaly/novelty detection objectives.

References

1. P. Afshar, S. Heidarian, F. Naderkhani, A. Oikonomou, K. N. Plataniotis and A. Mohammadi, Covid-caps: A capsule network-based framework for identification of covid-19 cases from x-ray images, *Pattern Recognition Letters* **138** (2020) 638–643.

2. S. Agarwal, N. S. Punn, S. K. Sonbhadra, P. Nagabhushan, K. Pandian and P. Saxena, Unleashing the power of disruptive and emerging technologies amid covid 2019: A detailed review, *arXiv preprint arXiv:2005.11507* (2020).

3. S. Alam, S. K. Sonbhadra, S. Agarwal and P. Nagabhushan, One-class support vector classifiers: A survey, *Knowledge-Based Systems* (2020) p. 105754.

4. I. D. Apostolopoulos and T. A. Mpesiana, Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks, *Physical and Engineering Sciences in Medicine* (2020) p. 1.

5. S. U. K. Bukhari, S. S. K. Bukhari, A. Syed and S. S. H. SHAH, The diagnostic evaluation of convolutional neural network (cnn) for the assessment of chest x-ray of patients infected with covid-19, *medRxiv* (2020).

6. E. Casiraghi, D. Malchiodi, G. Trucco, M. Frasca, L. Cappelletti, T. Fontana, A. A. Esposito, E. Avola, A. Jachetti, J. Reese *et al.*, Explainable machine learning for early assessment of covid-19 risk prediction in emergency departments, *IEEE Access* **8** (2020) 196299–196325.

7. CDC, Interim guidance for rapid antigen testing for sars-cov-2. `https://www.cdc.gov/coronavirus/2019-ncov/lab/resources/antigen-tests-guidelines.html`, (2020), [Online; accessed Oct 12, 2020].

8. G. J. Chowdary, N. S. Punn, S. K. Sonbhadra and S. Agarwal, Face mask detection using transfer learning of inceptionv3, in *International Conference on Big Data Analytics* (2020) pp. 81–90.

9. A. Christmann and I. Steinwart, How svms can estimate quantiles and the median, in *Advances in neural information processing systems* (2008) pp. 305–312.

10. J. P. Cohen, P. Morrison and L. Dao, Covid-19 image data collection, *arXiv preprint arXiv:2003.11597* (2020).

24   *S. K. Sonbhadra et al.*

11. R. Collobert, F. Sinz, J. Weston and L. Bottou, Trading convexity for scalability, in *Proceedings of the 23rd international conference on Machine learning* (2006) pp. 201–208.

12. S. Damiani, M. Fiorentino, A. De Palma, M. P. Foschini, T. Lazzarotto, L. Gabrielli, P. L. Viale, L. Attard, M. Riefolo and A. D'Errico, Pathological post mortem findings in lungs infected with sars-cov 2, *The Journal of pathology* (2020).

13. M. Farooq and A. Hafeez, Covid-resnet: A deep learning framework for screening of covid19 from radiographs, *arXiv preprint arXiv:2003.14395* (2020).

14. L. O. Hall, R. Paul, D. B. Goldgof and G. M. Goldgof, Finding covid-19 from chest x-rays using deep learning on a small dataset, *arXiv preprint arXiv:2004.02060* (2020).

15. S. Hasinoff, Computer vision: A reference guide, chap. photon, poisson noise, 608–610 (2014).

16. J. Hopkins, Covid-19 data repository by the center for systems science and engineering (csse) at johns hopkins university `https://github.com/CSSEGISandData/COVID-19`, (2020), [Online; accessed May 12, 2021].

17. X. Huang, L. Shi and J. A. Suykens, Support vector machine classifier with pinball loss, *IEEE transactions on pattern analysis and machine intelligence* **36**(5) (2013) 984–997.

18. J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, H. Marklund, B. Haghgoo, R. Ball, K. Shpanskaya *et al.*, Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison, in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33 (2019) pp. 590–597.

19. S. Jaeger, S. Candemir, S. Antani, Y.-X. J. Wáng, P.-X. Lu and G. Thoma, Two public chest x-ray datasets for computer-aided screening of pulmonary diseases, *Quantitative imaging in medicine and surgery* **4**(6) (2014) p. 475.

20. N. E. M. Khalifa, M. H. N. Taha, A. E. Hassanien and S. Elghamrawy, Detection of coronavirus (covid-19) associated pneumonia based on generative adversarial networks and a fine-tuned deep transfer learning model using chest x-ray dataset, *arXiv preprint arXiv:2004.01184* (2020).

21. S. S. Khan and M. G. Madden, A survey of recent trends in one class classification, in *Irish conference on artificial intelligence and cognitive science* (2009) pp. 188–197.

22. P. J. Kim, H. J. Chang and J. Y. Choi, Fast incremental learning for one-class support vector classifier using sample margin information, in *2008 19th International Conference on Pattern Recognition* (2008) pp. 1–4.

23. R. Koenker and K. F. Hallock, Quantile regression, *Journal of economic perspectives* **15**(4) (2001) 143–156.

24. A. Krizhevsky, I. Sutskever and G. E. Hinton, Imagenet classification with deep convolutional neural networks, in *Advances in neural information processing systems* (2012) pp. 1097–1105.

25. X. Ma, M. Liang, M. Ding, W. Liu, H. Ma, X. Zhou and H. Ren, Extracorporeal membrane oxygenation (ecmo) in critically ill patients with coronavirus disease 2019 (covid-19) pneumonia and acute respiratory distress syndrome (ards), *Medical science monitor: international medical journal of experimental and clinical research* **26** (2020) e925364–1.

26. A. Narin, C. Kaya and Z. Pamuk, Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks, *arXiv preprint arXiv:2003.10849* (2020).

27. L. Oakden-Rayner, Chexnet: an in-depth review. `https://lukeoakdenrayner.wordpress.com/2018/01/24/chexnet-an-in-depth-review/(2020)`, (2020), [Online; accessed Oct 02, 2020].

28. Y. Oh, S. Park and J. C. Ye, Deep learning covid-19 features on cxr using limited training data sets, *IEEE Transactions on Medical Imaging* (2020).

29. OpenCV, Image thresholding. `https://docs.opencv.org/master/d7/d4d/tutorial_py_thresholding.html`, (2020), [Online; accessed Oct 02, 2020].

30. G. Pang, C. Shen and A. van den Hengel, Deep anomaly detection with deviation networks, in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (2019) pp. 353–362.

31. N. S. Punn and S. Agarwal, Automated diagnosis of covid-19 with limited posteroanterior chest x-ray images using fine-tuned deep neural networks, *Applied Intelligence* (2020) 1–14.

32. N. S. Punn, S. K. Sonbhadra and S. Agarwal, Monitoring covid-19 social distancing with person detection and tracking via fine-tuned yolo v3 and deepsort techniques, *arXiv preprint arXiv:2005.01385* (2020).

33. A. Rizzi, C. Gatta and D. Marini, A new algorithm for unsupervised global and local color correction, *Pattern Recognition Letters* **24**(11) (2003) 1663–1677.

34. B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola and R. C. Williamson, Estimating the support of a high-dimensional distribution, *Neural computation* **13**(7) (2001) 1443–1471.

35. P. K. Sethy and S. K. Behera, Detection of coronavirus disease (covid-19) based on deep features, *Preprints* **2020030300** (2020) p. 2020.

36. S. K. Sonbhadra, S. Agarwal and P. Nagabhushan, Target specific mining of covid-19 scholarly articles using one-class approach, *Chaos, Solitons & Fractals* **140** (2020) p. 110155.

37. A. Stein, Pneumonia dataset annotation methods. rsna pneumonia detection challenge discussion, 2018. `https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/discussion/`, (2020), [Online; accessed Oct 02, 2020].

38. S. B. Stoecklin, P. Rolland, Y. Silue, A. Mailles, C. Campese, A. Simondon, M. Mechain, L. Meurice, M. Nguyen, C. Bassi *et al.*, First cases of coronavirus disease 2019 (covid-19) in france: surveillance, investigations and control measures, january 2020, *Eurosurveillance* **25**(6) (2020) p. 2000094.

39. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, Going deeper with convolutions, in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015) pp. 1–9.

40. M. Tan and Q. Le, Efficientnet: Rethinking model scaling for convolutional neural networks, in *International Conference on Machine Learning* (2019) pp. 6105–6114.

41. D. M. Tax and R. P. Duin, Support vector domain description, *Pattern recognition letters* **20**(11-13) (1999) 1191–1199.

42. L. Wang, Z. Q. Lin and A. Wong, Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images, *Scientific Reports* **10**(1) (2020) 1–12.

43. WHO, Who corona-viruses (covid-19) `https://www.who.int/emergencies/diseases/novel-corona-virus-2019,2020`, (2020), [Online; accessed May 02, 2020].

44. WHO, Who director-general's opening remarks at the media briefing on covid-19-11 march 2020. `https://www.who.int/dg/speeches/detail/`, (2020), [Online; accessed March 12, 2020].

45. Y. Xiao, W. Huangang and X. Wenli, Ramp loss based robust one-class svm, *Pattern Recognition Letters* **85** (2017) 15–20.

46. J. Zhang, Y. Xie, Y. Li, C. Shen and Y. Xia, Covid-19 screening on chest x-ray images using deep learning based anomaly detection, *arXiv preprint arXiv:2003.12338* (2020).