A two-tier feature selection method using Coalition game and Nystrom sampling for screening COVID-19 from chest X-Ray images

Pratik Bhowal¹ · Subhankar Sen² · Ram Sarkar³

Received: 19 March 2021 / Accepted: 31 August 2021 / Published online: 22 September 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

The world is still under the threat of different strains of the coronavirus and the pandemic situation is far from over. The method, that is widely used for the detection of COVID-19 is Reverse Transcription Polymerase chain reaction (RT-PCR), which is a time-consuming method and is prone to manual errors, and has poor precision. Although many nations across the globe have begun the mass immunization procedure, the COVID-19 vaccine will take a long time to reach everyone. The application of artificial intelligence (AI) and computer-aided diagnosis (CAD) has been used in the domain of medical imaging for a long period. It is quite evident that the use of CAD in the detection of COVID-19 is inevitable. The main objective of this paper is to use convolutional neural network (CNN) and a novel feature selection technique to analyze Chest X-Ray (CXR) images for the detection of COVID-19. We propose a novel two-tier feature selection method, which increases the accuracy of the overall classification model used for screening COVID-19 CXRs. Filter feature selection models are often more effective than wrapper methods as wrapper methods tend to be computationally more expensive and are not useful for large datasets dealing with a large number of features. However, most filter methods do not take into consideration how a group of features would work together, rather they just look at the features individually and decide on a score. We have used approximate Shapley value, a concept of Coalition game theory, to deal with this problem. Further, in the case of a large dataset, it is important to work with shorter embeddings of the features. We have used CUR decomposition and Nystrom sampling to further reduce the feature space. To check the efficacy of this two-tier feature selection method, we have applied it to the features extracted by three standard deep learning models, namely *VGG16*, *Xception* and *InceptionV3*, where the features have been extracted from the CXR images of COVID-19 datasets and we have found that the selection procedure works quite well for the features extracted by *Xception* and *InceptionV3*. The source code of this work is available at <https://github.com/subhankar01/covidfs-aihc>.

Keywords COVID-19 · Chest-X-Ray images · Deep learning · Feature selection · Coalition game theory · Nystrom sampling · CUR decomposition

Pratik Bhowal and Subhankar Sen contributed equally.

✉ Pratik Bhowal
pratikhhowal1999@gmail.com

Subhankar Sen
subhankarsen2001@gmail.com

Ram Sarkar
raamsarkar@gmail.com

¹ Department of Instrumentation and Electronics Engineering, Jadavpur University, Kolkata, India

² Department of Computer Science and Engineering, Manipal University Jaipur, Jaipur, India

³ Department of Computer Science and Engineering, Jadavpur University, Kolkata, India

1 Introduction

Coronavirus 2019 (COVID-19) pandemic, caused by an infectious virus named *Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)*, has led to unforeseen and challenging circumstances worldwide, putting an unprecedented standstill to the natural functioning of the human society, having devastating implications on the global economy, and has increased the overall mortality rate worldwide. Its initial cases were identified towards the end of 2019 (Synowiec et al. 2021). While most COVID-19 positive people experienced mild to moderate respiratory problems, some developed fatal lung infections. Typical indicators of the illness

include high fever, nausea, cough, exhaustion, and lack of taste and odor. Clinical characteristics differ in various cases, and some patients have an extremely communicable asymptomatic infection leading to uncontrollable dissemination of this viral pathogen worldwide (Yang et al. 2020).

Therefore, early detection of COVID-19 disease is of utmost importance to timely isolate the infected patients and thereby breaking the chain of infection. There is a pressing need for efficient screening and rapid medical response for affected patients. Reverse transcription polymerase chain reaction (RT-PCR) is a nuclear-derived laboratory technique used for detection and quantification of RNA virus infections and is by far the most extensively used clinical screening tool for COVID-19 identification using respiratory specimens. However, the RT-PCR experiment is a very time-consuming manual procedure that has often contributed to a great deal of subjectivity. There is a lack of research capability and expertise in several developing nations.

Secure and safe vaccination would be a game-changer. However, a handful number of reliable vaccines have been produced to date and it is estimated that it might probably take a very long time for the entire world to get immunized against the hazards of COVID-19. As a result, readily available radiological imaging comes into play as a viable alternative for performing diagnostic assessments and analysis of COVID-19 cases.

The diagnostic efficacy and detection of COVID-19 infections using radiological images such as Chest X-Ray (CXR), Computed Tomography (CT) scan, or Lung ultrasound (LUS) improved measurably with the emergence of artificial intelligence (AI) based technologies for Computer-aided diagnostic (CAD) applications, integrated with deep learning techniques. In the realms of end-to-end image processing and pattern recognition, deep learning algorithms have recently been shown to outperform conventional machine learning approaches (Ching et al. 2018; Iglovikov and Shvets 2018; Shen et al. 2017). Deep learning using convolutional neural network (CNN) is considered to be one of the robust and effective architectures for medical imaging processing, especially in classification and segmentation problems. In many cases, features are not directly fed to classifiers as they tend to have high dimensionality which gives rise to several problems like lower accuracy, higher training time, and overfitting. Feature selection is hence used as a preprocessing technique before features are fed to the classifier. Feature selection methods reduce the dimensionality of the features fed into the classifiers which in turn reduce training time, reduce overfitting, increase generalization and improve classification accuracy.

Feature selection algorithms are divided into three categories, namely filter methods, wrapper methods, and embedded methods (Guha et al. 2020; Ghosh et al. 2019b, a). In the filter method, the features are ranked

by a specific score, and this score is calculated by looking at how correlated the features are with each other and also how correlated they are with the true class labels. In the wrapper method, a search strategy is used to select the optimal feature subset. Feature selection methods in which the selection is a part of the optimization objective are known as an embedded methods. The main disadvantage of wrapper feature selection methods is that they are biased towards the classifier used for the selection procedure. Moreover, it has a very large search space, namely, 2^n where n is the number of features. Due to this, wrapper methods are less effective in large datasets with a large value of n . In the case of embedded models, we have a structured assumption of predictive tasks.

Filter methods, on the other hand, rely on the different statistical measures and are not dependent on specific classifier performance and thus there is no bias towards any classifier. Moreover, in practical situations, the complexity of the filter methods is much lower than wrapper methods. Besides, the classification accuracy of the filter method is almost equal to that of the wrapper methods in most practical situations. Thus in the case of high dimensional data, almost inevitably filter methods are used. However, many filter methods do not take into consideration the fact that even if a feature increases the classification accuracy it may not be compatible with other features. Thus a few features may be highly correlated with the true class labels and hence may individually be very important for classification, but together, they may not perform well. In other words, selecting those together may cause the classification accuracy to drop.

To deal with this problem, we have used *Coalition Game theory* along with *Information theory*. *Shapley Value*, a concept introduced in the Coalition Game theory, is used to calculate the contribution of a particular element (here feature) in the grand coalition formed. Hence, this method tries to analyze how compatible a specific feature is with respect to the other features that would be selected.

Another major problem in machine learning is the approximation of large matrices. The feature selection problem can be modeled as a similar problem with the samples acting as rows and the features as columns. An effective solution to this problem is *Nystrom sampling*, which generates a low-rank approximation of the original matrix (in this case the feature set) from a subset of its columns.

We have proposed a two-tier feature selection method in which we first use Coalition Game theory as a filter method to select features that are most compatible with each other and provide us with the maximum classification accuracy, and then use Nystrom Sampling to further reduce the number of columns or in this case the number of features.

1.1 Motivation and contributions

The use of deep CNN, transfer learning, and feature selection methods have achieved numerous breakthroughs and promising results in the field of image processing and computer vision for healthcare and medical diagnosis (Born et al. 2020a; Sahiner et al. 2019; Zahedi et al. 2018; Poplin et al. 2018; Adalarasan and Malathi 2018). To this end, the primary motivation of this paper is to propose an automated tool integrated with deep learning and feature selection approaches for efficient computerized detection and diagnosis of COVID-19. Feature selection has been used in many models where deep features have been extracted.

The main contributions of the present work are enlisted below:

1. We have applied deep feature extraction techniques from the CXR images of COVID-19 using fine-tuned CNN architectures, pre-trained on the ImageNet dataset for obtaining salient visual descriptors.
2. We have developed a novel two-tier feature selection technique, where, we have used Coalition Game Theory (approximate Shapley value) in tier one, and CUR decomposition and Nystrom Sampling in tier two. We are the first paper to combine Nystrom Sampling with Coalition Game theory. More importantly, since we cannot calculate the Shapley value for such large features, we introduced the Constrained Coalition Game which can be used to calculate the importance of features extracted by a CNN. To the best of our knowledge, this has been used first time for the purpose of feature selection.
3. The Constrained Coalition Game theory aimed at selecting features extracted by the CNN model has been introduced for the first time in this paper. Selecting features extracted by CNN cannot be done by Coalition Game theory and Shapley value due to the high time complexity of the Shapley value calculation procedure. We have tried to reduce the exponential time complexity to a linear one using the Constrained Coalition Game theory.

The rest of the paper is organized as follows. Section 2 gives an overview of the literature related to pulmonary disease detection with particular regard to COVID-19 disease. Section 3 provides a detailed description of our proposed approach. Section 4 presents the dataset description, experimental results obtained by employing the proposed feature selection on the features extracted from CNN models, and further discussion including comparison with the state-of-the-art in Sect. 4.4. Finally, the paper is concluded in Sect. 5 where future directions are provided.

2 Related work

Researchers have used many end-to-end machine learning and deep learning approaches to diagnose COVID-19 based on CXR images, CT scans, and LUS. This section discusses the newly developed CAD systems in the field of COVID-19 identification.

2.1 COVID-19 screening from CXR images

Turkoglu (2020) presented a COVID-19 diagnostic method called the COVIDetectionNet model. In this method, deep features were extracted using pre-trained AlexNet architecture from the CXR images, followed by the selection of the salient features from the obtained image descriptors using the Relief feature selection algorithm, followed by classification using the support vector machine (SVM) classifier. However, finding optimal parameters for the Relief algorithm and SVM classifier can be seen as a limitation that affected the performance of the above study. Sahlol et al. (2020) developed an enhanced hybrid classification method for COVID-19 image detection by developing a combination of pre-trained Inception model for deep feature extraction from CXRs accompanied by a swarm-based algorithm for selection of salient features, known as the Marine Predators Algorithm, enhanced by fractional-order calculus (FO) for selection of the most significant features.

Ucar and Korkmaz (2020) introduced a SqueezeNet CNN architecture with a deep Bayesian optimization additive, known as COVIDiagnosis-Net, for rapid identification and diagnosis of COVID-19 from CXR images. The fine-tuning of the SqueezeNet model performed on the ImageNet dataset, and hyperparameter tuning increases the robustness and efficiency of the proposed approach. Gianchandani et al. (2020) presented an ensemble of deep transfer learning models for COVID-19 detection using CXR images. The proposed framework was used for deep feature extraction from the input images. Due to the availability of a limited patient dataset, the training and learning capacity of the proposed models are greatly impacted which is the shortcoming of this proposed methodology.

Babukarthik et al. (2020) propounded a deep learning approach namely Genetic Deep Learning Convolutional Neural Network (GDCNN). It was trained from scratch for extracting features for classifying them between COVID-19 and normal images. The GDCNN architecture's mechanism was based on the genetic algorithm (GA) in which the CNN architecture gets refined using genetic operations. In another study conducted by Elaziz et al. (2020), a machine learning method was proposed to classify the CXR images into two classes (COVID-19 and non-COVID). The features were extracted from the CXR images using new Fractional

Multichannel Exponent Moments (FrMEMs), followed by a modified Manta-Ray Foraging Optimization based on differential evolution used for feature selection.

Singh et al. (2021b) initially used CNN variants such as GA-CNN and PSO-CNN for COVID-19 classification from CXR images. However, since techniques like GA and Particle Swarm Optimization (PSO) suffer from the limitation of poor and pre-mature convergence issues, the authors employed a novel meta-heuristic algorithm, Multi-objective Adaptive Differential Evolution (MADE) for overcoming the convergence issue and for hyperparameter tuning. Panetta et al. (2021) presented a shape-dependent Fibonacci -p patterns-based feature extractor for distilling out the intricate textural features from CXR images. This methodology has the inherent advantage of being computationally inexpensive and is tolerant to noise and illumination present in the CXR images. The authors of Kaur et al. (2021) presented a meta-heuristic-based deep learning model in which uses a modified AlexNet as a deep feature extractor and classification of input CXR images. Strength Pareto Evolutionary Algorithm-II (SPEA-II) was used for model optimization and hyperparameter tuning in their proposed methodology. Dey et al. (2021) developed a Choquet Fuzzy Integral-based ensemble model for COVID-19 screening from CXR images. Pre-trained, transfer learning DCNN models such as VGG19, DenseNet121 and InceptionV3 were used as feature extractors for distilling out discriminating features from the CXR images followed by classification using a classifier. The classifier outputs of the three models were then aggregated using Choquet Integral-based ensemble. The authors of Karbhari et al. (2021) presented the approach of Auxiliary Classifier Generative Adversarial Network (ACGAN) which was used for the synthesis of COVID-19 positive and normal CXR images. State-of-the-art CNN architectures were trained on synthetic images and used as deep feature extractors. This was followed by the application of the Harmony Search (HS) algorithm for feature selection which facilitated dimensionality reduction of the extracted feature vector.

2.2 COVID-19 screening from CT scans

El-Kenawy et al. (2020) presented two optimization algorithms for feature space reduction and COVID-19 detection from radiological CT scans. The proposed framework was comprised of three cascaded phases. At first, the features were derived from the CT scans by employing the AlexNet architecture. Further, the authors implemented a feature selection algorithm known as the Guided Whale Optimization Algorithm (Guided WOA) based on Stochastic Fractal Search (SFS). Finally, Particle Swarm Optimization (PSO)-based Guided WOA was used as a voting classifier for the aggregation of different decision scores obtained using four

classifiers, namely Neural Network (NN), Decision Tree (DT), k-nearest neighbors (KNN), and SVM.

Li et al. (2020) proposed a 3D deep learning-based architecture, coined as COVID-19 detection neural network (COVNet), was implemented for extraction of features from chest CT scans. Local 2D and global 3D visual features were extracted using the proposed methodology. However, several limitations and shortcomings were highlighted in the proposed study. The paucity of clinical evidence of the etiology for cases of viral pneumonia led to the inability in selecting other viral cases of pneumonia for the comparative analysis. Furthermore, the proposed method wasn't able to address the problem of categorizing lung disease in terms of differing severity levels. Sen et al. (2021), the authors employed an efficient COVID-19 screening model which was divided into two modules. In the first module, CNN was used for feature extraction from the CT scans. In the second module, a bi-staged feature selection technique, a guided approach using mutual information and Relief-F followed by the Dragonfly algorithm, was used for selecting the top k salient features from the CT images.

Zhang et al. (2020) performed a research analysis proposing an 8-layer enhanced CNN model with stochastic pooling and hyperparameter optimization. Stochastic pooling was adopted to replace normal average pooling and max pooling. However, the dataset used by this method is relatively small and some new networks are not tested, which are outlined as limitations by the authors. Chattopadhyay et al. (2021) performed deep residual feature extraction from CT scans using pre-trained ResNet18. The authors coined a meta-heuristic feature selection technique, known as Cluster-based Golden Ratio based Optimizer (CGRO) which employed clustering-based population generation for prevention of premature algorithm convergence. Singh et al. (2021a) used an ensemble of deep transfer learning models, such as ResNet152V2, DenseNet201, and VGG16, to address the sensitivity issue which is associated with RT-PCR for COVID-19 detection. The paper used chest CT scans for evaluation of the proposed ensemble model for a 4-class classification setting (COVID-19, pneumonia, tuberculosis, and healthy cases). Wang et al. (2020c) revamped the original COVID-Net by improving network architecture for COVID-19 detection from CT scan images. The authors also introduced contrastive cross-site learning and a learning rate scheduling strategy based on cosine annealing which enhanced the joint training protocol and facilitated domain invariance enhancement, remarkably reducing the inter-site data heterogeneity. Garain et al. (2021) developed a three-layer deep convolution spiking neural network (DCSNN) for COVID-19 detection from CT scans. The input images are convolved and processed by the means of Gabor filters. This is followed by the generation of a wave of spikes with the aid of intensity-to-latency encoding. The spike-wave reaches the

penultimate layer after they are propagated through a series of convolutional and pooling layers of the CNN architecture. Finally, salient features are extracted from the images which are fed as input into a classifier for decision generation.

2.3 COVID-19 screening from LUS

The work reported in Roy et al. (2020) elucidates a deep learning architecture was employed by Roy et al. that forecasted the score correlated with single LUS images and identifies regions with weakly monitored pathological characteristics. They have created a lightweight approach that focused on aggregating frame-level predictions and estimating the score associated with the LUS video sequences. Born et al. (2020b) developed a frame-based DCNN for the classification of COVID-19 using LUS video clips. The authors of the paper used class activation maps (CAMs) for visualizing the spatial-temporal location of pulmonary bio-markers, which were eventually tested for human-in-the-loop (HITL) scenarios in a blindfold trial with medical professionals.

In all of the above methods used for CXR image classification, which involve feature selection, features have been selected using Evolutionary Algorithms or swarm-based intelligence. These methods, however, are supervised and use classifiers which makes them computationally more expensive. Our method on the other hand proposes an unsupervised method for feature selection.

3 Proposed work

3.1 Method overview

First, the CXR images are downsized to 512×512 pixels using bicubic interpolation, a 2D method using cubic splines to sharpen and enhance the resolution of digital images. To deal with the imbalance problem of the dataset, offline image augmentation techniques are used. Fine-tuned, well-established DCNN architectures, containing millions of parameters and pre-trained over the ImageNet dataset namely VGG16, Xception, and InceptionV3 are used for the extraction of deep salient features from each input CXR image. By employing deep feature extraction, we can reduce the likeliness of overfitting and there is a drastic decline in the number of trainable parameters. This is followed by the process of selecting a subset of relevant features with the aid of a novel two-tier feature selection technique. In the first tier, we have employed Coalition Game theory using approximated Shapley Values which are followed by CUR decomposition and Nystrom sampling for the second tier. These salient image descriptors are then fed as input into a Multi-layered perceptron (MLP) classifier with softmax output for the 3-classification problem (COVID-19 vs Pneumonia vs

Normal). Figure 1 illustrates the overall block diagram of the proposed method.

3.2 Data preprocessing and augmentation

The CXR images of different resolutions are downsampled to 512×512 pixels using bicubic interpolation before they are fed as input into the DCNN models for deep feature extraction. We then perform color space conversion of the CXR images from RGB to grayscale followed by the application of Gaussian mixture model-based adaptive Wiener filter for denoising the CXR images (Tabuchi et al. 2007).

Further, image augmentation techniques are applied to the pre-processed images to build a generalized classification model by incorporating the random variability in the images. Ideally, we can augment the images in a way such that the key and discriminating features required for making predictions are retained, but the image pixel orientations are rearranged or distorted to add some variation to the input images. This enhances the robustness of the deep learning models by increasing the classification complexity and hence problems such as overfitting can be curbed. Various image augmentation techniques such as image rotation, horizontal flip, shearing, horizontal and vertical image translation are used.

3.3 Deep feature extraction and fine-tuning

Feature extraction is an essential component in the field of computer vision, image processing, and pattern recognition systems. Deep feature extraction is the process of exploiting DCNN models in which the input image is allowed to propagate along with the depth of the DCNN models followed by taking the outputs as features from some pre-specified layers of the DCNN model. The key purpose of deep feature extraction is to capture discriminating information from the image and depict the information in a lower-dimensional space. We use the deep feature extraction technique from CXR images using fine-tuned DCNN models. Salient features are extracted from the feature maps corresponding to specific convolution layers using the *global average pooling* operation.

The deep features are then concatenated into a single feature vector, known as image descriptor, as represented in Table 1. We have employed the Keract library (Remy 2020) for studying the activation maps corresponding to each layer in the DCNN architectures and then the suitable convolution layers for feature extraction are determined. This technique facilitates the interpretability and debugging of deep black-box neural networks. Fig. 2 depicts the activation maps generated corresponding to some unique, hand-picked convolutional layers of the three DCNN models used for deep feature extraction.

Fig. 1 Block diagram of the proposed COVID-19 screening method from CXR images

VGG16, Xception, and InceptionV3, pre-trained on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) dataset, are the DCNN models that have been used in this work. This is because robust deep learning models can be built with comparatively less training data with the aid of transfer learning without training the neural network from scratch, thereby optimizing the classification performance over the unseen test images.

The *Global average pooling* operation computes the average output of all the feature maps in the previous layer and hence acts as a structural regularizer that explicitly enforces feature maps to be confidence maps of categories. Furthermore, it summarizes the spatial features, making it more robust for obtaining the salient spatial representation of the input images (Lin et al. 2013). For instance, using the global average pooling operation a 3D $[H, W, C]$ tensor is converted into a 3D tensor of shape $[1, 1, C]$ by computing the average over the $H \times W$ slices, which is then reshaped into a 1D vector

of length equal to the number of channels (C). In Fig. 3, VGG16 is shown to understand how to derive discriminating features from input CXR images.

The fine-tuning strategy of the pre-trained DCNN models is performed which facilitates faster convergence of gradient descent, drastically reduces the number of trainable parameters, and enhances the model's ability to learn domain-specific patterns. With the aid of fine-tuning, optimal experimental results were obtained by keeping the ImageNet weights of VGG16, Xception, InceptionV3 in a non-trainable state up to 17th, 175th, and 675th convolutional layers respectively. This enables us to retain the generic low-level visual features in the CXRs such as edges and blobs.

3.4 Proposed feature selection method

In the proposed method, we use the two-tier feature selection method. Initially, we use a *constrained coalition game*

Fig. 2 Investigation of deep features by employing activation map visualization of the three DCNN models used

Fig. 3 Schematic diagram of VGG16 architecture for deep feature extraction from the CXR images

Table 1 Details of regions of feature extraction from the given DCNN models

DCNN models	Layer no.	Feature dimension
VGG16	5,13,17	1152
Xception	27, 37, 126	2992
InceptionV3	11, 18, 28, 51, 74, 101, 120,152, 184, 216, 249, 263, 294	2320

The third column depicts the dimension of the deep feature vector extracted from the CXR images

theory, to select features from those set of features extracted by the deep learning models. The features selected by this method are then further reduced in dimension using *CUR decomposition* and the *Nystrom sampling*.

The coalition game is used to understand the contribution that a feature has on the classification process. Coalition game takes into consideration not only the contribution of the feature but also the contribution of the feature when considered with other features. To capture this concept, we use *Shapley value*, a concept used in the coalition game theory, which is used to calculate the payoff for each element (in this case features) when it is assumed that a grand coalition of elements has been formed. The formula for calculating the Shapley value is as follows

$$\varphi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(n - |S| - 1)!}{n!} (v(S \cup i) - v(S)). \quad (1)$$

The term $(v(S \cup i) - v(S))$ is referred to as the *marginal contribution*. It indicates the contribution that the element i has on the subset of elements S . As seen above, it is calculated as the difference between the values of the characteristic function (v) of the subset which contains the element i and the value of the characteristic function (v) of the subset which does not contain the element i but contains all other elements that are present in the last subset S . If the marginal contribution is higher it indicates that adding i to the already existing subset inflates the value of v a great deal, thereby inflating the gain of that subset.

We use *Information theory* to evaluate the *Marginal contribution*. Features that are very correlated to each other and whose correlation with the true label reduces when they are considered together are known as *Redundant features*. If two features extracted are not related to each other and if their aptness to the true label does not improve when they are taken into consideration together, they are referred to as *Independent features*. Two features are referred to as *Interdependent features* if their aptness to the true label improves when they are taken into consideration together instead of being examined separately. *Mutual information*

and *Conditional mutual information* are employed by us to comprehend how interdependent, redundant, or independent a feature is with the set of features that were previously selected.

Let P be the subset of features that have been previously selected and let f_x be the feature that we want to add to this existing subset. Let $I_{mu}(P; class)$ represents the mutual information between the true label of the class and the set of features P and $I_{mu}(f_x; class|P)$ represents the mutual information between the true label of the class and the feature f_x given the set of features P . Then, the following formulas hold.

$$I_{mu}(f_x; class|P) < I_{mu}(P; class) \text{ for redundancy,} \quad (2)$$

$$I_{mu}(f_x; class|P) > I_{mu}(P; class) \text{ for interdependency,} \quad (3)$$

$$I_{mu}(f_x; class|P) = I_{mu}(P; class) \text{ for independence.} \quad (4)$$

We calculate the marginal contribution by taking the difference between the mutual information and the conditional mutual information. This difference gives us an intuitive idea about the contribution of the element (feature) f_x in the subset of elements (features) P . Therefore, we encapsulate the information about the true label that we gain after the inclusion of f_x to the already selected subset of features P . Therefore, the equation becomes,

$$\text{marginal contribution} = I_{mu}(f_x; class|P) - I_{mu}(P; class) \quad (5)$$

We do not calculate the *Marginal contribution* using Eq. 5 as it is extremely difficult and in many cases impossible to compute. We are using an approximate formula to calculate the same. It is evident that the marginal contribution of f_x would be low if the mutual information between the true label and the f_x is low, while the mutual information between P and f_x is high. We devise Eq. 6 taking this logic into consideration. We assume that p is the cardinality of the set P and the term I_{mu} are as defined before.

$$\text{marginal contribution} = I_{mu}(f_x; class) - \frac{1}{p} \sum_{f_p \in P} I(f_x, f_p). \quad (6)$$

Even this approximate formula gives rise to high complexity as for each of the features we need to calculate its Shapley value, and for calculating this, we need to consider all possible subsets of the features that can be formed by the rest of the features. It is not hard to understand that this is computationally very expensive with time complexity of $2^{n-1}n$.

To deal with this complex problem, we introduce the concept of *constrained Shapley value*. It is based on the assumption that when an image is fed to the CNN model, only a specific region of the image is taken into consideration by each node of the CNN, and the subsequent nodes take into consideration only a specific region of the feature

maps generated by the layer just before it. When our features are extracted, finally these are flattened out. We explain this hypothesis with the help of Fig. 4.

In Fig. 4 (a simplified demo-network to explain the hypothesis), we see that the bigger portion of the image is converted into a smaller area using the Convolution layer. This process is repeated a number of times in the CNN model and finally, in the flatten layer, a number of the features are combined to form a node. Thus, it is clear that a portion of the image is represented by a few nodes in the flattened layer. Again, since we know that portions of the image that far away have a little spatial relationship among each other, we may say that nodes in the flattened layer which are far away from each other have very little influence on each other in terms of the learning process of the CNN models. In the case of our features, we concatenate different flattened layers obtained after flattening different feature maps as described before. Thus in our case, the features far away from each other in the feature set have even less influence on each other. Hence, it is safe for us to assume that the feature for which we calculate the Shapley value makes a significant contribution to the subsets of features (i.e., has high marginal contribution value for that subset) that lie within a certain neighborhood of the feature.

Experimentally, we find that taking four features from the left and four features from the right gives us adequate results while keeping the complexity under check. Thus, the complexity decreases from $2^{n-1}n$, which is exponential to 2^8n , which is linear.

Top N features having the highest score (i.e., Shapley values) are selected and sent for the second tier of the proposed feature selection process where we use Nystrom sampling and

CUR decomposition. The value of N has been determined experimentally.

Let us assume that n be the number of samples present in the dataset and N is the number of features selected per sample by the constrained coalition game theory. We desire to reduce the number of selected features to p where $p < N$. In doing so, we first select p samples from the set of n samples and then Nystrom sampling. We use CUR decomposition for selecting the p samples. Since, we select p samples, after CUR decomposition, the dimension of the C matrix is $n * p$, the dimension of R matrix is $p * N$, while the dimension of the U matrix is $p * p$. We now use the R matrix for the Nystrom sampling, since the matrix R contains p samples from the dataset.

For Nystrom sampling, we first construct two similarity matrices, one of dimension $p * p$ which contains the similarity information among the p selected samples (in sample points) in the dataset while the other similarity matrix has the dimension $n * p$ which contains the similarity information among the p selected samples and the n samples originally present in the dataset. We represent the first similarity matrix by S_{in} while we represent the second similarity matrix by S_a .

S_{in} is then diagonalized using the Eq. 7. This is known as sample embedding.

$$S_{in} = Q\Sigma^2Q^T. \quad (7)$$

Let the number of non-zero eigenvectors be equal to $k < p$ and hence the matrix Q becomes Q_k and the matrix Σ becomes equal to Σ_k . Time and space complexity required for the evaluation of these two matrices are respectively $O(p^2k)$ and $O(p^2)$.

Let U_k be the k dimensional representation of all the n samples and thus it has dimension equal to $n * k$. Since we

Fig. 4 Architecture of simple CNN model to explain the proposed hypothesis. This diagram shows how a considerable portion of the image is converted to a few nodes (features) after being flattened and hence during the calculation of the Shapley value of a feature, we consider only a few features in the neighborhood of the feature

already know the embeddings of the sample points, we use the properties of the similarity matrix to derive the embeddings of the other points. This can be done using Eq. 8.

$$S_a = U_k(Q_k \Sigma_k)^T. \quad (8)$$

We can calculate U_k using Eq. 9, which is derived from Eq. 8 by post-multiplying each side of the equation by $Q_k \Sigma_k^{-1}$ and due to the property that $Q_k^T Q_k = I_k$.

$$U_k = S_a Q_k \Sigma_k^{-1}. \quad (9)$$

4 Experimental results and discussion

4.1 Dataset description

A dataset, known as *Novel COVID-19 Chestxray Repository*, by combining various publicly available CXR image repositories is prepared. Three different datasets obtained from Github and Kaggle, created by the authors of other research studies in this field, are combined to form the current dataset. Frontal and lateral CXR radiographic images are used in this study since these views of radiography are widely used by radiologists for clinical examination. Table 3 tabulates the dataset composition of the proposed Novel COVID-19 Chestxray Repository. In the subsequent section, we have outlined the details of the parent image repositories which are used for the combination.

- **COVID-19 radiography database** A team of researchers at the University of Dhaka and Qatar University had partnered with medical professionals from Pakistan and Malaysia to compile a CXR image database known as the COVID-19 Radiography Database (Chowdhury et al. 2020a). This first release of this dataset comprises a total

of 219 COVID-19 positives, 1345 viral pneumonia, and 1341 normal radiographic CXR images, which are updated periodically with the emergence of new COVID-19 cases across the globe. The Kaggle database is publicly available at (Chowdhury et al. 2020b).

- **COVID-Chestxray set** A public image repository on Github (Cohen et al. 2020), consisting of both CT scans and digital CXR images, was created by Joseph Paul Cohen and Paul Morrison, and Lan Dao. Data were primarily obtained from longitudinal cohort studies of pediatric cases in Guangzhou Women and Children's Medical Center in China. With the aid of metadata information provided along with the dataset, 521 COVID-19 positives, 239 viral and bacterial pneumonia, and 218 normal radiographic CXR images were obtained.
- **Actualmed COVID chest X-Ray dataset** 12 COVID-19 positives and 80 normal radiographic CXR images were found in the initial release of the Actualmed COVID chest X-Ray dataset. The image archive is publicly available on Github (Wang et al. 2020b). This dataset was formed by researchers working at DarwinAI Corp., Canada and Vision, and Image Processing Research Group, University of Waterloo, Canada. It consists of radiographs of the CXR data modality. This dataset is open-source and is being updated periodically with the contributions of several researchers working in the field of COVID-19 screening.

The combined image repository consists of digital CXR image files of COVID-19 positive, pneumonia, and normal (healthy) classes, with a total of 752, 1584, and 1639 images respectively. The dataset is split into training, validation, and test sets consisting of a total of 3085, 771, and 120 (40 randomly selected images taken from each class) CXR images respectively. The train validation split in the ratio of 4:1. Table 2 displays the class distribution of the

Fig. 5 **a** COVID-19 positive CXR images, **b** pneumonia infected CXR, and **c** normal CXR images present in Novel COVID-19 Chestxray Repository

Table 2 Class distribution of the proposed Novel COVID-19 Chestxray Database for performance evaluation of the proposed method

	COVID-19	Pneumonia	Normal	Total
Number of images per class				
Training data (80%)	570	1235	1280	3085
Validation data (20%)	142	309	319	771
Test data	40	40	40	120

proposed COVID-19 dataset used for this analysis. The dataset is available publicly on Github and Kaggle. CXR images in each class are shown in Fig. 5.

4.2 Experimental setup

Our experiment is implemented in Python using the Keras package with Tensorflow used as the deep learning framework backend and run on Google Colaboratory having

the following system specifications: Nvidia Tesla T4 with 1.59 GHz GPU Memory Clock, 13 GB GPU memory, and 12.72 GB RAM.

Adam optimizer, with a learning rate of 0.001, and a hyperparameter of β_1 and β_2 set of 0.6 and 0.8 respectively, are used to train the MLP with the derived image descriptors. For model tuning and optimization, the learning rate and hyperparameter values are experimentally inferred to be the most optimal values obtained using the *grid search* technique. The batch size is set equal to 32, training of the DCNN models is performed up to 1000 epochs. For all DCNNs, weights are initialized from weights learned on ImageNet.

The obtained matrix U_k comprising the final feature embedding of each sample is eventually fed to an MLP consisting of a fully connected (FC) layer of 512 neurons (with activation of the rectified linear unit or ReLU) accompanied by a dropout rate of 0.5. The dropout is vital to curtail the overfitting of the neural network. There is eventually the output layer of three neurons (with softmax activation) for the 3-class CXR image classification problem under consideration.

Table 3 Statistics of the CXR images used from different publicly available repositories to form the Novel COVID-19 Chestxray Repository

Dataset	COVID-19	Pneumonia	Normal
COVID Chestxray set (Cohen et al. 2020)	521	239	218
COVID-19 Radiography Database (Chowdhury et al. 2020b)	219	1345	1341
Actualmed COVID chestxray dataset (Wang et al. 2020b)	12	0	80
Total number of CXRs	752	1584	1639

Fig. 6 Accuracy after choosing the different values of N while performing feature selection using Coalition game for the features extracted using the three CNN models

Table 4 Accuracy (in %) after choosing fixed values of N and varying the values of k and calculating the overall accuracy for each of the cases for VGG16

Values of N	Values of k	Accuracy after the entire feature selection
1000	750	93.33
	500	92.50
	250	81.60
	100	79.17
750	500	90.00
	250	79.17
	100	77.50
500	250	76.66
	100	75.00
250	100	75.00

Table 5 Accuracy (in %) after choosing fixed values of N and varying the values of k and calculating the overall accuracy for each of the cases for Xception

Values of N	Values of k	Accuracy after two-tire feature selection
1000	750	94.17
	500	93.33
	250	87.50
	100	76.66
750	500	90.00
	250	87.50
	100	83.33
500	250	76.66
	100	75.80
250	100	74.16

4.3 Results

In this paper, we have performed a 3-class classification problem, where the images of CXRs are classified as normal lungs, normal pneumonia infected lungs, and COVID-19 infected lungs.

We have used three CNN models to extract the features. Feature selection and embedding methods have been applied to the extracted features and finally, an MLP has been used to classify the features and in turn the images. In this section, we report the results of the feature selection technique on the three sets of extracted features, i.e., features extracted by the three CNN models.

In Fig. 6, we have reported the experimental outcomes conducted for the selection of the value N in a bar graph plot. From the experiments, it is clear that the best results are obtained when $N = 500$ for *InceptionV3* and $N = 750$

Table 6 Accuracy (in %) after choosing fixed values of N and varying the values of k and calculating the overall accuracy for each of the cases for InceptionV3

Values of N	Values of k	Accuracy after two-tire feature selection
1000	750	93.33
	500	91.66
	250	85.00
	100	81.60
750	500	92.50
	250	87.50
	100	79.17
500	250	85.00
	100	76.66
250	100	75.00

for both *VGG16* and *Xception*. However, it is to be noted that this is a two-tire feature selection method and hence, the results do not depend only on N , but on the combined value of N and k . Thus, a specific value of N , can give good results in tire one and yet may not yield good results for any value of k in tire two.

To demonstrate this set of experimentation, we show various results in Tables 4, 5 and 6. In Table 4, we show the results of the experiments for *VGG16* for different values of N and k , while in Tables 5 and 6, we show the results of the experiments for *Xception* and *InceptionV3* respectively.

In Table 7, we depict how the tire two of the proposed feature selection method, i.e., the Nystrom sampling, improves the overall result. In this table, we compare the

Table 7 Accuracy (in %) of the models proposed before and after Nystrom sampling

CNN model used for feature extraction	Accuracy before Nystrom sampling	Accuracy after Nystrom sampling
VGG16	91.66	93.33
Xception	92.50	94.17
InceptionV3	93.33	93.33

Table 8 Accuracy (in %) of the models proposed before and after the combined feature selection method of Coalition game theory and Nystrom sampling

CNN model used for feature extraction	Accuracy before feature selection	Accuracy after feature selection
VGG16	95.00	93.33
Xception	92.50	94.17
InceptionV3	92.50	93.33

Fig. 7 ROC plots of the three DCNN models employed before and after applying proposed two-tier feature selection

best performances for all the three models after the first tire of feature selection (due to different values of N) and after

the second tire of feature selection (due to different values of N and k).

In Table 8, we compare the accuracy before and after the two-tier feature selection for each of the three models.

From Tables 4, 5 and 6, we can see that the best results are achieved when we set $N = 1000$ and $k = 750$. We can see from Fig. 6 and Table that a greater accuracy after tire one of the feature selection does not guarantee high accuracy after the two tiers of feature selection. For example, in the case of *InceptionV3*, the best accuracy after tire one of the feature selection is when we select $N = 750$ but after tier two of the feature selection process are over, we find that the performance is best when we select $N = 1000$ and $k = 750$. Moreover, from Table 8, we can see that the feature selection method works for both *Xception* and *InceptionV3*. However, in the case of *VGG16*, the feature selection method does not work properly and the accuracy drops.

A Receiver Operating Characteristic (ROC) is a graphical plot which depicts the relationship between sensitivity and

Fig. 8 ROC plots for the three classes, the micro-averaged and macro-averaged after classification after the feature selection has been performed on the three DCNN models used

specificity. It is used as an evaluation metric for measuring the classification performance obtained at various threshold settings. The ROC curve is a graph with:

- x-axis representing false positive rate (FPR) = $FP/(FP + TN)$,
- y-axis representing true positive rate (TPR) = $TP/(TP + FN)$,

where FP, TN, TP, FN denote the number of false positives, true negatives, true positives, and false negatives respectively.

Figure 7 illustrates the ROC Curves of the three pre-trained DCNN models employed before and after applying proposed two-tier feature selection. Figure 8 depicts the ROC curves for each of the three classes and the micro-average and macro-average of the three classes for *VGG16*, *Xception* and *InceptionV3* respectively. Figure 9 depicts the confusion matrix for each of the three models, namely, *VGG16*, *Xception* and *InceptionV3* respectively.

4.4 Comparison with other methods

Since the classification problem, considered here, there is no standard dataset that has been used by most of the past researchers, it is difficult to directly compare the proposed method with other methods. However, we have chosen one such dataset which has been used in two research works in the recent past. For the comparison, with the papers (Oh et al. 2020; Wang et al. 2020a) (COVIDNet), we have used the COVIDx dataset. In Table 9 we have depicted the comparison with the said methods.

Wang et al. (2020a) proposed a novel, lightweight DCNN architecture, COVIDNet, for COVID-19 screening from CXR images of the COVIDx dataset. The authors of the paper used projection-expansion-projection-extension (PEPX) design which improves the representational capacity, thereby drastically reducing the computational complexity. However, the absence of feature normalization layers in COVIDNet's architecture culminates in a significant variance of the learned representations across different layers of the model, which can be seen as a limitation of their proposed methodology. Oh et al. (2020) developed a deep learning approach inspired by statistical analysis of the potential

Fig. 9 Confusion matrices representing the classification performance after the proposed two-tier feature selection has been performed on the three DCNN models employed

imaging biomarkers of CXR images. After performing standard image preprocessing techniques, the CXR images are fed into a segmentation network (FC-DenseNet103) which detects the pulmonary regions and outputs lung contours. The segmented lung regions are then passed through a CNN classifier which is trained for outputting patch-level predictions. The final decisions are then obtained using majority voting.

5 Conclusion and future directions

In the method that we have proposed in this paper, for COVID-19 screening from CXRs, we have extracted deep features from the images and have proposed a two-tier feature selection method before classification. In the first tier, we have used coalition game theory, whereas in the second tier we have used Nystrom sampling. The Shapley value that we have used here takes care of not only how the features perform individually but also how well they perform together as a combination. As the calculation of Shapley value is an extremely computationally expensive task, we have used an approximation to make it less computationally expensive. In the second tier, Nystrom sampling helps us transform our feature embedding into a lower-dimensional space while retaining the important features.

While calculating the Shapley value, we assume that a grand coalition always forms. Thus when we look at the contribution of the individual features, we also look at its interaction with other features which are not selected at the end. Thus a feature may perform poorly with some of the features which are not even selected at the end and may perform well with some of the features which are part of the final subset of features. But even after this, the feature may be rejected because of its low Shapley value which

Table 9 Comparison of the proposed method with the methods reported in Wang et al. (2020a) and Oh et al. (2020) on the COVIDx dataset

Performance metric	Wang et al. (2020a)	Oh et al. (2020)	Proposed method
Accuracy (%)	93.3	–	93.45
Normal sensitivity	0.950	0.900	0.92
Pneumonia sensitivity	0.940	0.930	0.935
COVID 19 sensitivity	0.910	1.000	0.98
Normal (precision)	0.913	0.957	0.920
Pneumonia (precision)	0.938	0.903	0.950
COVID 19 (precision)	0.889	0.769	0.87

indicates low contribution, without taking into consideration that its contribution is lower for features that are not even selected and is higher for features that are selected. Thus in this case Shapley's value makes a mistake while selecting features.

Besides this, we intend on working with other COVID-19 X-Ray and CT scan datasets. We also plan to apply the proposed feature selection method in other domains where high dimensional features and huge datasets are used like in Microarray datasets.

Funding All the authors declare that they have not received any kind of funding from any agencies.

Data availability The Novel COVID-19 Chestxray Repository can be accessed at <https://www.kaggle.com/subhankarsen/novel-covid19-chestxray-repository>.

Code availability The python code implementation is publicly available at <https://github.com/subhankar01/covidfs-aihc>.

Declarations

Conflict of interest All the authors declare that they have no conflict of interest or competing interests.

References

- Adalarasan R, Malathi R (2018) Automatic detection of blood vessels in digital retinal images using soft computing technique. *Mater Today Proc* 5(1):1950–1959
- Babukarthik R, Adiga VAK, Sambasivam G, Chandramohan D, Amudhavel J (2020) Prediction of covid-19 using genetic deep learning convolutional neural network (gcdnn). *IEEE Access* 8:177647–177666
- Born J, Brändle G, Cossio M, Disdier M, Goulet J, Roulin J, Wiedemann N (2020a) POCOVID-net: automatic detection of COVID-19 from a new lung ultrasound imaging dataset (pocus). [arXiv:2004.12084](https://arxiv.org/abs/2004.12084)
- Born J, Wiedemann N, Brändle G, Buhre C, Rieck B, Borgwardt K (2020b) Accelerating COVID-19 differential diagnosis with explainable ultrasound image analysis. [arXiv:2009.06116](https://arxiv.org/abs/2009.06116)
- Chattopadhyay S, Dey A, Singh PK, Geem ZW, Sarkar R (2021) COVID-19 detection by optimizing deep residual features with improved clustering-based golden ratio optimizer. *Diagnostics* 11(2):315
- Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, Ferrero E, Agapow PM, Zietz M, Hoffman MM, et al. (2018) Opportunities and obstacles for deep learning in biology and medicine. *J R Soc Interface* 15(141):20170387
- Chowdhury ME, Rahman T, Khandakar A, Mazhar R, Kadir MA, Mahbub ZB, Islam KR, Khan MS, Iqbal A, Al-Emadi N et al (2020a) Can ai help in screening viral and COVID-19 pneumonia? [arXiv:2003.13145](https://arxiv.org/abs/2003.13145)
- Chowdhury ME, Rahman T, Khandakar A, Mazhar R, Kadir MA, Mahbub ZB, Islam KR, Khan MS, Iqbal A, Al-Emadi N, Reaz MB

- (2020b) COVID-19 radiography database. <https://www.kaggle.com/tawsifurrahman/covid19-radiography-database>
- Cohen JP, Morrison P, Dao L, Roth K, Duong TQ, Ghassemi M (2020) COVID-19 image data collection: prospective predictions are the future. *arXiv:2006.11988*. <https://github.com/ieee8023/covid-chestxray-dataset>
- Dey S, Bhattacharya R, Malakar S, Mirjalili S, Sarkar R (2021) Choquet fuzzy integral-based classifier ensemble technique for covid-19 detection. *Comput Biol Med*. <https://doi.org/10.1016/j.compbiomed>
- El-Kenawy ESM, Ibrahim A, Mirjalili S, Eid MM, Hussein SE (2020) Novel feature selection and voting classifier algorithms for COVID-19 classification in CT images. *IEEE Access* 8:179317–179335
- Elaziz MA, Hosny KM, Salah A, Darwish MM, Lu S, Sahlol AT (2020) New machine learning method for image-based diagnosis of COVID-19. *PLoS One* 15(6):e0235187
- Garain A, Basu A, Giampaolo F, Velasquez JD, Sarkar R (2021) Detection of covid-19 from ct scan images: a spiking neural network-based approach. *Neural Comput Appl* 1–14
- Ghosh M, Guha R, Sarkar R, Abraham A (2019a) A wrapper-filter feature selection technique based on ant colony optimization. *Neural Comput Appl* 1–19
- Ghosh M, Kundu T, Ghosh D, Sarkar R (2019b) Feature selection for facial emotion recognition using late hill-climbing based memetic algorithm. *Multimedia Tools Appl* 78(18):25753–25779
- Gianchandani N, Jaiswal A, Singh D, Kumar V, Kaur M (2020) Rapid covid-19 diagnosis using ensemble deep transfer learning models from chest radiographic images. *J ambient intell humanized comput* 1–13
- Guha R, Ghosh M, Mutsuddi S, Sarkar R, Mirjalili S (2020) Embedded chaotic whale survival algorithm for filter-wrapper feature selection. *Soft Comput* 24(17):12821–12843
- Iglovikov V, Shvets A (2018) Terausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation. *arXiv:1801.05746*
- Karbhari Y, Basu A, Geem ZW, Han GT, Sarkar R (2021) Generation of synthetic chest X-Ray images and detection of COVID-19: a deep learning based approach. *Diagnostics* 11(5):895
- Kaur M, Kumar V, Yadav V, Singh D, Kumar N, Das NN (2021) Metaheuristic-based deep COVID-19 screening model from chest X-ray images. *J Healthc Eng*
- Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, Bai J, Lu Y, Fang Z, Song Q, Cao K (2020) Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. *Radiology*
- Lin M, Chen Q, Yan S (2013) Network in network. *arXiv:1312.4400*
- Oh Y, Park S, Ye JC (2020) Deep learning covid-19 features on cxr using limited training data sets. *IEEE Trans Med Imaging* 39(8):2688–2700
- Panetta K, Sanghavi F, Agaian S, Madan N (2021) Automated Detection of COVID-19 Cases on Radiographs using Shape-Dependent Fibonacci-p Patterns. *IEEE J Biomed Health Inform*
- Poplin R, Varadarajan AV, Blumer K, Liu Y, McConnell MV, Corrado GS, Peng L, Webster DR (2018) Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng* 2(3):158–164
- Remy P (2020) Keract: a library for visualizing activations and gradients. <https://github.com/philipperemy/keract>
- Roy S, Menapace W, Oei S, Luijten B, Fini E, Saltori C, Huijben I, Chennakeshava N, Mento F, Sentelli A, Peschiera E (2020) Deep learning for classification and localization of COVID-19 markers in point-of-care lung ultrasound. *IEEE Trans Med Imaging* 39(8):2676–2687
- Sahiner B, Pezeshk A, Hadjiiski LM, Wang X, Drukker K, Cha KH, Summers RM, Giger ML (2019) Deep learning in medical imaging and radiation therapy. *Med Phys* 46(1):e1–e36
- Sahlol AT, Yousri D, Ewees AA, Al-Qaness MA, Damasevicius R, Abd Elaziz M (2020) COVID-19 image classification using deep features and fractional-order marine predators algorithm. *Sci R* 10(1):1–15
- Sen S, Saha S, Chatterjee S, Mirjalili S, Sarkar R (2021) A bi-stage feature selection approach for COVID-19 prediction using chest CT images. *Appl Intell* 1–6
- Shen D, Wu G, Suk HI (2017) Deep learning in medical image analysis. *Annu rev biomed eng* 19:221–248
- Singh D, Kumar V, Kaur M (2021a) Densely connected convolutional networks-based COVID-19 screening model. *Appl Intell* 51(5):3044–3051
- Singh D, Kumar V, Yadav V, Kaur M (2021b) Deep neural network-based screening model for covid-19-infected patients using chest x-ray images. *Inter J Pattern Recognit Artif Intell* 35(03):2151004
- Synowiec A, Szczepański A, Barreto-Duran E, Lie LK, Pyrc K (2021) Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2): a systemic infection. *Clin Microbiol Rev* 34(2):e00133–e00120
- Tabuchi M, Yamane N, Morikawa Y (2007) Adaptive wiener filter based on gaussian mixture model for denoising chest x-Ray CT image. In: *SICE Annual Conference 2007*. IEEE, pp 682–689
- Turkoglu M (2020) COVIDetectionNet: COVID-19 diagnosis system based on X-ray images using features selected from pre-learned deep features ensemble. *Appl Intell* 51(3):1213–1226
- Ucar F, Korkmaz D (2020) COVIDiagnosis-Net: Deep Bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images. *Med Hypotheses* 140:109761
- Wang L, Lin ZQ, Wong A (2020a) COVID-net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-Ray images. *Sci Rep* 10(1):1–12
- Wang L, Wong A, Lin ZQ, McInnis P, Chung A, Gunraj H (2020b) Actualmed COVID-19 chest X-Ray dataset initiative. <https://github.com/agchung/Actualmed-COVID-chestxray-dataset>
- Wang Z, Liu Q, Dou Q (2020c) Contrastive cross-site learning with redesigned net for COVID-19 CT classification. *IEEE J Biomed Health Inform* 24(10):2806–2813
- Yang Y, Yang M, Shen C, Wang F, Yuan J, Li J, Zhang M, Wang Z, Xing L, Wei J et al (2020) Laboratory diagnosis and monitoring the viral shedding of 2019-NCOV infections. *MedRxiv*
- Zahedi A, On V, Phandthong R, Chaili A, Remark G, Bhanu B, Talbot P (2018) Deep analysis of mitochondria and cell health using machine learning. *Sci Rep* 8(1):1–15
- Zhang YD, Nayak DR, Zhang X, Wang SH (2020) Diagnosis of secondary pulmonary tuberculosis by an eight-layer improved convolutional neural network with stochastic pooling and hyperparameter optimization. *J Ambient Intell Human Comput* 1–18

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.