Quantitative Biology 2022, 10(2): 208–220
https://doi.org/10.15302/J-QB-021-0278

RESEARCH ARTICLE

COVIDX: Computer-aided diagnosis of COVID-19 and its severity prediction with raw digital chest X-ray scans

Wajid Arshad Abbasi[1,*], Syed Ali Abbas[1], Saiqa Andleeb[2], Maryum Bibi[1], Fiaz Majeed[3], Abdul Jaleel[4], Muhammad Naveed Akhtar[5]

[1] Computational Biology and Data Analysis Lab, Department of Computer Sciences & Information Technology, King Abdullah Campus, University of Azad Jammu & Kashmir, Muzaffarabad, AJ&K 13100, Pakistan
[2] Biotechnology Lab, Department of Zoology, King Abdullah Campus, University of Azad Jammu & Kashmir, Muzaffarabad, AJ&K 13100, Pakistan
[3] Department of Software Engineering, University of Gujrat, Gujrat 50700, Pakistan
[4] Department of Computer Science, (RCET), UET, Lahore 54000, Pakistan
[5] Computational and Internet Services Division, Pakistan Institute of Engineering and Applied Sciences (PIEAS), Islamabad 44000, Pakistan
* Correspondence: wajidarshad@gmail.com

Received May 30, 2021; Revised July 27, 2021; Accepted August 25, 2021

***Background*: Coronavirus disease (COVID-19) is a contagious infection caused by severe acute respiratory syndrome coronavirus-2 (SARS-COV-2) and it has infected and killed millions of people across the globe.**
***Objective*: In the absence or inadequate provision of therapeutic treatments of COVID-19 and the limited convenience of diagnostic techniques, there is a necessity for some alternate spontaneous screening systems that can easily be used by the physicians to rapidly recognize and isolate the infected patients to circumvent onward surge. A chest X-ray (CXR) image can effortlessly be used as a substitute modality to diagnose the COVID-19.**
***Method*: In this study, we present an automatic COVID-19 diagnostic and severity prediction system (COVIDX) that uses deep feature maps of CXR images along with classical machine learning algorithms to identify COVID-19 and forecast its severity. The proposed system uses a three-phase classification approach (healthy vs unhealthy, COVID-19 vs pneumonia, and COVID-19 severity) using different conventional supervised classification algorithms.**
***Results*: We evaluated COVIDX through 10-fold cross-validation, by using an external validation dataset, and also in a real setting by involving an experienced radiologist. In all the adopted evaluation settings, COVIDX showed strong generalization power and outperforms all the prevailing state-of-the-art methods designed for this purpose.**
***Conclusions*: Our proposed method (COVIDX), with vivid performance in COVID-19 diagnosis and its severity prediction, can be used as an aiding tool for clinical physicians and radiologists in the diagnosis and follow-up studies of COVID-19 infected patients.**

Availability: We made COVIDX easily accessible through a cloud-based webserver and python code available at the site of google and the website of Github.

Keywords: coronavirus; COVID-19; radiology; machine learning; chest X-ray; contagious infection

Author summary: Coronavirus disease (COVID-19) is a contagious infection that has killed masses across the world. Due to the lack of specific therapeutics for the treatment of COVID-19, timely diagnosis of the infection is critical to circumvent its further surge by recommending isolation or quarantine. The normal clinical screening test to diagnose COVID-19 is complex, manual, costly, and laborious. To combat this, we have proposed a machine learning based computer-aided diagnosis (CAD) system called COVIDX (COVID-19 Detection using X-ray images) to identify coronavirus and predict its severity using digital X-ray scans. The performance of the proposed system has been verified rigorously through different measures. The well verified solution is publically available through a cloud based webserver for free use.

INTRODUCTION

Coronavirus disease (COVID-19) is a contagious infection caused by severe acute respiratory syndrome coronavirus-2 (SARS-COV-2) and its transmission is also possible from asymptotic patients while the incubation period [1,2]. This infectious pandemic has killed masses across the world [3]. The World Health Organization (WHO) in its reports has already acknowledged this epidemic as a global health catastrophe [4]. According to medical experts, this disease primarily infects the human breathing system triggering severe pneumonia by showing symptoms of dry cough, breathing problems, fever, fatigue, and lung failure, etc. [1,5]. Right now, the world is curiously waiting for some therapeutic treatments to prevent this lethal infection.

Due to the lack of specific therapeutics for the treatment of COVID-19, timely diagnosis of the infection is critical to circumvent its further surge by recommending isolation or quarantine. Presently, there exist two sorts of tests to identify COVID-19: antibody tests and diagnostic tests [6]. The test is based on antibodies and gazes for antibodies formed in the infected body in reaction to the infection. Whereas an antibody test cannot effectively be adopted to identify COVID-19 as after infection, antibodies require weeks to develop and could stay in one's blood a bit extended after retrieval [6]. Meanwhile, diagnostic tests such as reverse transcription-polymerase chain reaction (RT-PCR) can reliably be used to detect active COVID-19 infection. Mostly, RT-PCR is the common analytic exercise to identify viral infections [7]. However, the normal clinical RT-PCR screening test to diagnose COVID-19 is complex, manual, costly, and laborious [8]. Furthermore, an inadequate convenience of test kits and domain specialists additionally obstruct the case. Keeping in view all these restraints of fundamental diagnostic procedures and a swift flow of diseased patients, there is an immediate necessity for some substitute and easy screening systems that can effectively be to rapidly diagnose and quarantine the diseased patients.

X-ray imaging is quite a common noninvasive indicative method that supports physicians to analyze and treat several infections. A chest X-ray (CXR) is normally taken to assess the medical fitness of the lungs, heart, and chest wall. The CXR has also been playing a critical role in the pilot investigation of various respiratory irregularities [9,10]. In this context, a CXR image can also be used as an alternative modality to detect and diagnose the COVID-19 infection. CXR images are normally interpreted by expert radiologists. However, some studies indicated that CXR images of patients, infected by COVID-19, showed various features [5,8,11]. Therefore, the elucidation of these CXR images with slight deviations is quite puzzling manually. Moreover, the current massive increase of diseased patients makes the task of completing timely diagnostic even more challenging for the area specialists [9,12]. To combat this situation, some reliably accurate computer-aided diagnosis (CAD) systems are the need of the time.

In CAD using X-ray images, machine learning has already been applied successfully in many clinical and radiological studies to describe various characteristics of radio imaging [13–15]. Also, to identify coronavirus infection using CXR scans, a plethora of machine learning-based techniques have been proposed in the literature with the varying amount and sources of training datasets (for further detail readers can consult the "Related works" section). Most of these previously published studies are based on a deep learning approach using convolution neural network (CNN) [16]. However, techniques based on the deep learning paradigm, to apply properly, normally require a massive amount of time and data which is sometimes not feasible [17]. Meanwhile, some studies adopted transfer learning where fine-tune the final classification layer also involves a huge amount of data to compensate high-dimensional features maps obtained through pre-trained models [17–19]. To combat these issues, a study has been proposed using a classical machine learning method built on majority voting and handcrafted texture descriptors of X-ray images [9]. However, there still exists a possible scope of performance enhancement even after using majority voting and image augmentation because handcrafted texture descriptors are unable to represent subtle variations of X-ray images appropriately [9,20]. In the meanwhile, to the best of our knowledge, nearly all the related proposed studies focused only on the diagnoses of COVID-19 and we didn't find any method to predict its severity. Likewise, all the proposed methods in the literature only presented the technical details of the adopted methodology in their method without giving any user-friendly interface for its proper use by the practitioners in the field. To compensate for these loopholes or shortcomings, we have proposed in this study a novel predictor called COVIDX (COVID-19 detection using X-ray images) to identify coronavirus and predict its severity. We have designed COVIDX by engaging a merger of transfer and classical machine learning. An easily accessible webserver of COVIDX has also been made freely available along with open source python code.

RELATED WORKS

While going through the literature on machine learning techniques used for the detection of COVID-19 with digital chest X-ray images, we found a plethora of machine learning-based techniques in the literature with the varying amount and sources of training datasets [8,9,12,17–19,21–31]. Generally, all of these previously available methods can broadly be clustered into three clusters: (1) Methods employing deep learning tactics; (2) Methods employing transfer-learning approaches via fine-tuning; (3) Methods employing classical-learning paradigm with different handcrafted texture-based features.

Mostly, deep learning-based proposed machine learning methods in the literature to diagnose COVID-19 implemented CNN [16] centered frameworks [8,12,21–31]. However, deep learning-based machine learning techniques, to apply properly, normally require a massive amount of time and data which is sometimes not feasible [17].

Spotting data scarceness, some studies have also been proposed to diagnose coronavirus infection by employing transfer learning-based approaches where available pre-trained models have been used to get features by removing the top classification layer [17–19]. Khan *et al.* proposed a method based on Xception architecture called CoroNet to detect COVID-19 infection from chest X-ray images with an overall accuracy of 89.6% [18]. Minaee *et al*. used different CNN-based pre-trained models such as ResNet18, ResNet50, SqueezeNet, and DenseNet-121 to diagnose COVID-19 by employing a transfer learning approach and achieved 90% specificity [19]. Similarly, Wang *et al*. also used the transfer learning approach by removing the classification layer on the top of the pre-trained model, adding a custom classification layer as the top layer, and then by freezing the convolutional layers in front of the framework to train only the customized classification layer [17]. The model proposed by Wang *et al.* achieved an accuracy of 96.75% [17]. These approaches to diagnose COVID-19 through CXR by employing transfer-learning are still not perfectly precise because in transfer-learning, to fine-tune the customized classification layer also demands an enormous amount of data because of high-dimensional features maps obtained through available standard pre-trained models.

To compensate for the problems of data scarceness in deep learning-based approaches to diagnose COVID-19, Chandar *et al*. proposed a classical machine learning-based method by using majority voting and handcrafted texture descriptors of X-ray images [9]. However, even after using majority voting and image augmentation, the technique presented by Chandra *et al.* still offered room to improve generalization accuracy because of limitations in representing subtle distinctions in CXR scans using texture-based handcrafted feature space [20]. Besides, nearly all the related studies proposed in the literature had been designed only to focus on the diagnoses of COVID-19. Whereas, COVID-19 severity prediction still requires research attention.

RESULTS AND DISCUSSIONS

In this work, we have proposed a machine learning-based model for automatic coronavirus diagnosis and its severity prediction. We have divided the COVID-19 diagnostic task into three sub-tasks based on the available data: (1) classification of X-ray images as healthy vs unhealthy, (2) classification of X-ray images as COVID-19 vs pneumonia, and (3) COVID-19 severity prediction. For this purpose, we have used various machine learning algorithms and deep feature maps. In what follows we present results showing the prediction performance of our proposed method by using cross-validation and an external validation dataset. While presenting the predictive performance of the proposed system, we have also compared its performance with two previously published state-of-the-art approaches: (1) deep learning-based approach [17] and (2) shallow approach with handcrafted texture-based features [9].

Predictive performance for the classification of healthy versus unhealthy X-ray scans

The results for the classification of healthy versus unhealthy X-ray scans with the adopted experimental procedure along with 10-fold CV and external validation dataset are shown in Tables 1 and 2 and Fig. 1A, B. We got a maximum F1 score of 0.99 using 10-fold CV. The values of the area under the ROC curve and the area under the PR curve are 0.99 and 0.99 under this setting respectively, with the DenseNet121 feature map and SVC (Table 1). The score of 0.99 for F1, the area under the PR curve, and the area under the ROC curve show excellent sensitivity, which is important because we want as minimum missed diagnosis rate of COVID-19 as possible. Similarly, improved precision reported

Table 1 Predictive performance for the classification of healthy vs un-healthy X-ray scans across different machine learning techniques and feature representations using 10-fold CV

Feature map	SVC			RFC			XGBC		
	ROC	PR	F1	ROC	PR	F1	ROC	PR	F1
Resnet50	0.98±0.01	0.98±0.01	0.98	0.99±0.03	0.98±0.08	0.96	0.98±0.03	0.98±0.08	0.98
Xception	0.99±0.01	0.98±0.01	0.99	0.99±0.01	0.99±0.01	0.97	0.99±0.01	0.99±0.01	0.98
InceptionV3	0.98±0.01	0.99±0.01	0.97	0.99±0.01	0.99±0.02	0.97	0.99±0.02	0.99±0.01	0.97
VGG16	0.99±0.01	0.99±0.01	0.99	0.98±0.01	0.97±0.01	0.98	**0.99±0.01**	**0.99±0.01**	**0.98**
NASNetLarge	0.98±0.01	0.98±0.01	0.96	0.99±0.01	0.99±0.01	0.97	0.98±0.03	0.98±0.01	0.96
DenseNet121	**0.99±0.01**	**0.99±0.01**	**0.99**	**0.99±0.01**	**0.99±0.01**	**0.98**	0.96±0.03	0.96±0.01	0.97

SVC (support vector classifier), RF (random forest classifier), XGBC (XGBoost classifier), ROC (area under the ROC curve), PR (area under the precision-recall curve), F1 (F1 score). Bold-faced values specify the top performance for each of the models.

Table 2 Predictive performance for the classification of healthy vs un-healthy X-ray scans across different machine learning techniques and feature representations on an external validation dataset

Feature map	SVC			RFC			XGBC		
	ROC	PR	F1	ROC	PR	F1	ROC	PR	F1
Resnet50	0.97	0.97	0.91	0.98	0.97	0.91	0.97	0.96	0.93
Xception	0.98	0.97	0.91	0.97	0.97	0.91	0.97	0.97	0.93
InceptionV3	0.96	0.95	0.89	0.97	0.98	0.88	0.97	0.98	0.90
VGG16	0.97	0.97	0.91	0.98	0.97	0.91	0.98	0.97	0.91
NASNetLarge	0.96	0.95	0.87	0.96	0.96	0.85	0.96	0.96	0.88
DenseNet121	**0.99**	**0.98**	**0.98**	**0.98**	**0.97**	**0.94**	**0.99**	**0.98**	**0.96**

SVC (support vector classifier), RF (random forest classifier), XGBC (XGBoost classifier), ROC (area under the ROC curve), PR (area under the precision-recall curve), F1 (F1 score). Bold-faced values specify the top performance for each of the models.

through these performance metrics implies excellent performance also in classifying healthy cases. These results on 10-fold CV show an improved performance of our proposed method in comparison to the state-of-the-art deep learning-based method proposed by Wang *et al.* [17] with an F1 score of 0.96 ($p < 0.05$). We have achieved this performance improvement due to efficiently large dimensions' feature space handling ability of classical machine learning models in comparison to deep architectures in the presence of fewer examples. Meanwhile, if we compare the performance of our proposed method with the state-of-the-art shallow approach using handcrafted texture-based features under 10-fold CV, we find a comparable performance over the task [9].

Similarly, we have also re-confirmed the robustness of generalization ability of our proposed system over the classification task of healthy versus unhealthy X-ray scans using an external validation dataset to mitigate the possible bias performance improvement under 10-fold CV with hyperparameter tuning using the same training set. Using an external validation dataset for the re-evaluation of our proposed solution, we got an F1 score of 0.98. The values of the area under the ROC curve and the area under the PR curve are 0.99 and 0.98 under this setting respectively, with the DenseNet121 feature map and SVC (Table 2, Figure 1A, B). These metrics, using a validation dataset, show a robust and consistently outstanding generalization power of our proposed methodology in classifying healthy and unhealthy cases. These results show an improved performance of our proposed method in comparison to the state-of-the-art deep learning-based method proposed by Wang *et al.* [17] with an F1 score of 0.88 and a shallow learning-based method proposed by Chandra *et al.* [9] with an F1-score of 0.96 and the area under the ROC curve of 0.93 by using SVM (RBF kernel) ($p < 0.05$). This performance improvement in comparison to the method proposed by Chandra *et al.* is due to the use of deep feature maps instead of handcrafted texture-based features [20]. The significantly improved performance of SVM in comparison to XGBC and RFC is achieved due to its capacity to treat well the high-dimensional feature space with fewer instances as in our case. The consistent higher performance of feature maps extracted through the DenseNet121 pre-trained model is attributed towards fewer parameters of the pre-trained model. Moreover, we have attained better performance by using deep feature maps in comparison to handcrafted features as reported by Chandra *et al.* [9], in their study.

Predictive performance for the classification of COVID-19 versus pneumonia X-ray scans

The results for the classification of COVID-19 versus

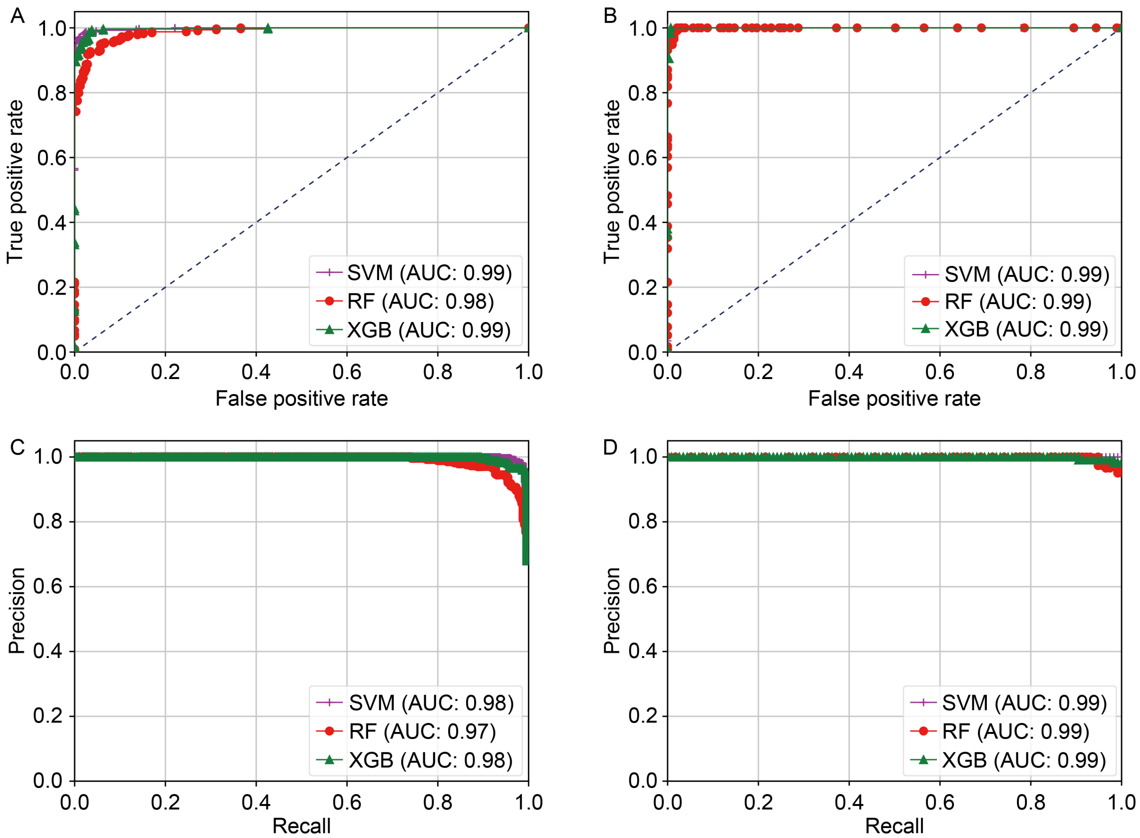

Figure 1. ROC and PR curves showing predictive performance of COVIDX for the classification of CXR scans (healthy vs unhealthy and COVID-19 vs pneumonia) across different classifiers (SVC, RF, XGB) and DenseNet feature map on an external validation dataset. (A, B) Healthy vs unhealthy. (C, D) COVID-19 vs pneumonia.

pneumonia X-ray scans with the adopted experimental procedure along with 10-fold CV and external validation dataset are shown in Tables 3 and 4 and Fig. 1C, D). We got a maximum F1 score of 0.99 using 10-fold CV. The values of the area under the ROC curve and the area under the PR curve are 0.99 and 0.99 under this setting respectively, with the DenseNet121 feature map and SVC (Table 3). These scores of reported accuracy metrics show an ideal sensitivity (classifying COVID-19 correctly) and specificity (classifying pneumonia correctly) of the proposed machine learning technique. This is very important as we do not want to isolate a pneumonia patient without medication while suspecting COVID-19 and we also do not want to give unnecessary medication to a COVID-19 patient without advising quarantine while suspecting pneumonia. These results across 10-fold CV show an improved performance of our proposed method in comparison to the state-of-the-art method proposed by Chandra *et al*. with a maximum F1-score of 0.84 and the area under the ROC curve of 0.85 with SVM (linear kernel) [9] ($p < 0.05$).

Meanwhile, to further confirm the robustness and generalization power of our proposed technique over the classification task of COVID-19 vs pneumonia X-ray images, we have also reported the performance metrics on an external validation dataset. While using an external validation dataset for the evaluation of our machine learning models trained on the training set we observed a similar trend of performance with an F1-score of 0.99 along with 0.99, and 0.99 as the area under the ROC curve, and the area under the PR curve, respectively with Support Vector Classifier and DenseNet121 feature map (Table 4, Fig. 1C, D). These metrics, using a validation dataset, show a robust and consistently outstanding generalization of our proposed methodology in classifying COVID-19 and pneumonia cases. These results also show a better performance of our proposed method in comparison to the state-of-the-art shallow method proposed by Chandra *et al*., with an F1-score of 0.91 and the area under the ROC curve of 0.91 even after using majority voting [9] ($p < 0.05$). The significantly improved performance of SVM in comparison to XGBC and RFC is achieved due to its capacity to treat well the high-dimensional feature space with fewer instances as in our case. Moreover, we have attained better performance by using deep feature maps

Table 3 Predictive performance for the classification of COVID-19 vs pneumonia X-ray scans across different machine learning techniques and feature representations using 10-fold CV

Feature map	SVC			RFC			XGBC		
	ROC	PR	F1	ROC	PR	F1	ROC	PR	F1
Resnet50	0.98±0.01	0.97±0.01	0.97	0.99±0.03	0.98±0.08	0.96	0.98±0.03	0.98±0.08	0.98
Xception	0.98±0.01	0.97±0.01	0.98	0.99±0.01	0.99±0.01	0.97	0.99±0.01	0.99±0.01	0.98
InceptionV3	0.98±0.01	0.97±0.01	0.97	0.99±0.01	0.99±0.02	0.97	0.99±0.02	0.99±0.01	0.97
VGG16	0.98±0.01	0.97±0.01	0.98	0.98±0.01	0.97±0.01	0.98	0.99±0.01	0.99±0.01	0.98
NASNetLarge	0.97±0.01	0.97±0.01	0.96	0.99±0.01	0.99±0.01	0.97	0.98±0.03	0.98±0.01	0.96
DenseNet121	**0.99±0.01**	**0.99±0.01**	**0.99**	**0.99±0.01**	**0.99±0.01**	**0.98**	**0.99±0.03**	**0.99±0.01**	**0.99**

SVC (support vector classifier), RF (random forest classifier), XGBC (XGBoost classifier), ROC (area under the ROC curve), PR (area under the precision-recall curve), F1 (F1 score). Bold-faced values specify the top performance for each of the models.

Table 4 Predictive performance for the classification of COVID-19 vs pneumonia X-ray scans across different machine learning techniques and feature representations on an external validation dataset

Feature map	SVC			RFC			XGBC		
	ROC	PR	F1	ROC	PR	F1	ROC	PR	F1
Resnet50	0.99	0.99	0.98	0.99	0.98	0.96	0.99	0.98	0.98
Xception	0.99	0.99	0.98	0.97	0.97	0.95	0.97	0.97	0.95
InceptionV3	0.99	0.99	0.98	0.97	0.98	0.96	0.97	0.98	0.96
VGG16	0.99	0.99	0.99	0.98	0.97	0.91	0.98	0.97	0.94
NASNetLarge	0.98	0.99	0.98	0.96	0.96	0.90	0.96	0.96	0.93
DenseNet121	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.98**	**0.99**	**0.99**	**0.99**

SVC (support vector classifier), RF (random forest classifier), XGBC (XGBoost classifier), ROC (Area under the ROC curve), PR (area under the precision-recall curve), F1 (F1 score). Bold-faced values specify the top performance for each of the models.

in comparison to handcrafted features as reported by Chandra *et al.* [9].

Performance over COVID-19 severity prediction

In addition to COVID-19 diagnosis, we have also trained machine learning modes to predict the severity of COVID-19 infection to further assist the decisions related to intensive health care. The results for the classification of severe vs mild COVID-19 infections using X-ray scans with the adopted experimental procedure along 10-fold CV are shown in Table 5. We got a maximum F1 score of 0.90 using 10-fold CV. The values of the area under the ROC curve and the area under the PR curve are 0.96 and 0.90 under this setting respectively, with the DenseNet121 feature map and SVC (Table 5). These results show a reasonable performance of COVIDX over the severity prediction of coronavirus infection. The F1 score of 0.90 along with the area under the ROC curve of 0.96 shows reasonable sensitivity (classifying severe COVID-19 patients correctly) of our proposed model. This is what we want in this case as we do not want to miss severe COVID-19 patients who require intensive health care. Meanwhile, reasonable specificity (classifying mild COVID-19 patients correctly) helps in reducing the unnecessary use of limited intensive health care facilities. These results show a reasonable performance of COVIDX over the severity prediction of COVID-19 infection.

Performance of the proposed computer-aided COVID-19 diagnostic system in a real setting

We have also evaluated the performance of COVIDX in a real setting using its webserver under the observation of a qualified radiologist at a local hospital at Muzaffarabad, AK, Pakistan. We performed this performance evaluation both for coronavirus diagnosis and its severity prediction. For COVID-19 diagnosis, 90 X-ray images (anonymous) belonging to different classes (30 COVID-19 infected, 30 pneumonia infected, and 30 healthy persons) have been used. The results of this evaluation in the form of a confusion matrix are shown in Fig. 2A. Our proposed system (COVIDX) classified correctly 26 out of 30 provided X-ray scans of COVID-19 infected patients, 3 as pneumonia infection and 1 as healthy (Fig. 2A). Likewise, for the provided X-ray scans of pneumonia infected patients, our proposed system correctly classifies 25 out of 30 scans, 4 as COVID and 1 as healthy (Fig. 2A). Furthermore, on health X-ray scans, our system correctly classified 27 out of 30 X-ray scans as healthy, 2 as pneumonia

Table 5 Predictive performance for the severity prediction of corono virus across different machine learning techniques and feature representations using 10-fold CV

Feature map	SVC			RFC			XGBC		
	ROC	PR	F1	ROC	PR	F1	ROC	PR	F1
Resnet50	0.92±0.12	0.82±0.28	0.86	0.85±0.23	0.73±0.31	0.80	**0.96±0.11**	**0.89±0.17**	**0.90**
Xception	0.87±0.14	0.74±0.28	0.82	0.81±0.25	0.67±0.29	0.78	0.86±0.28	0.78±0.29	0.82
InceptionV3	0.86±0.13	0.68±0.29	0.74	0.73±0.30	0.58±0.30	0.75	0.82±0.26	0.76±0.29	0.78
VGG16	0.90±0.24	0.88±0.26	0.83	0.81±0.24	0.70±0.24	0.77	0.82±0.26	0.72±0.31	0.80
NASNetLarge	0.80±0.31	0.76±0.30	0.76	0.80±0.26	0.65±0.30	0.76	0.79±0.26	0.68±0.31	0.75
DenseNet121	**0.96±0.17**	**0.90±0.19**	**0.90**	**0.90±0.19**	**0.78±0.26**	**0.84**	0.90±0.20	0.83±0.26	0.90

SVC (support vector classifier), RF (random forest classifier), XGBC (XGBoost classifier), ROC (area under the ROC curve), PR (area under the precision-recall curve), F1 (F1 score). Bold-faced values specify the top performance for each of the models.

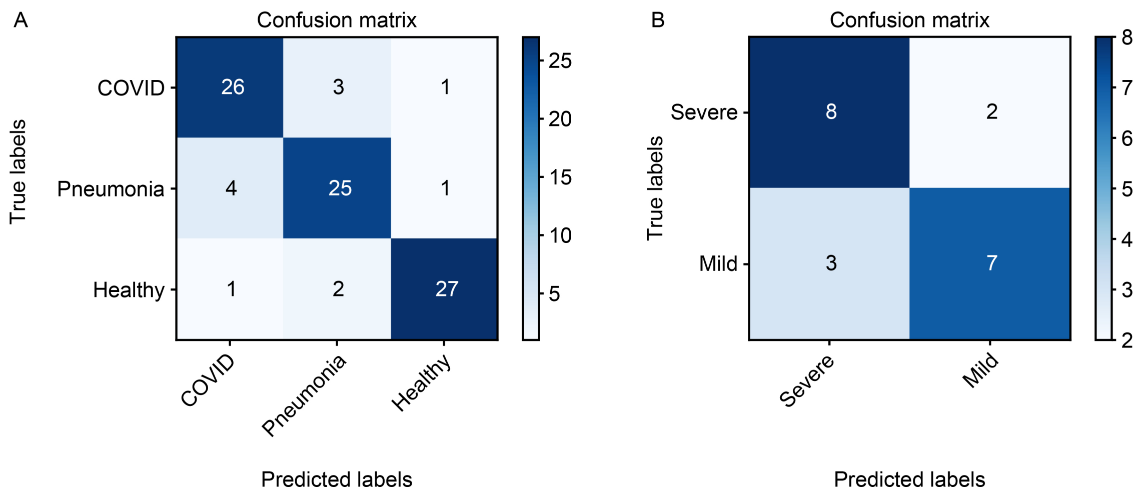

Figure 2. Confusion matrix depicting COVIDX real performance under the observation of a qualified radiologist. (A) COVID-19 diagnosis. (B) COVID-19 severity prediction.

infection and 1 as COVID-19 (Fig. 2A). These results validate the possible use of our proposed approach through its webserver in a real setting for coronavirus infection diagnosis.

Similarly, to predict COVID-19 severity, we have used 20 X-ray scans (anonymous) belonging to different classes (10 severe COVID-19 infections, 10 mild infections). The results of this evaluation in the form of a confusion matrix are shown in Fig. 2 B. Our proposed system (COVIDX) classified properly 8 out of 10 provided X-ray scans of severe COVID-19 infected patients and 2 as mild (Fig. 2B). Likewise, for the used X-ray scans of mildly infected patients, our system categorized properly 7 out of 10 scans and 3 as severe (Fig. 2B). These results show that COVIDX can also efficiently be used while taking intensive health care decisions for COVID-19 patients.

DISCUSSION

A computer-aided automatic system to diagnose COVID-19 and predict its severity is important due to the lack of specific therapeutics for its treatment. The effective use of X-ray imaging along with machine learning techniques in many other clinical and radiological studies dictates this imaging modality to be an important area to be explored for the diagnosis of COVID-19 infection. Several machine learning-based computer-aided methods, for this purpose, have already been proposed in the literature that use digital X-ray images. The majority of these published methods use deep learning from scratch or transfer learning by fine-tuning the classification layer. These approaches based on the deep or transfer learning paradigm face data scarceness problems. Meanwhile, already published studies based on classical machine learning techniques along with handcrafted features did not show an up-to-the-mark performance. Additionally, almost all the published techniques focused only on COVID-19 diagnosis, but its severity prediction is still awaiting attention. To handle these issues of data scarceness and to get the finest possible outcome of both worlds, we have proposed a fresh technique that combines transfer and classical machine learning to diagnose COVID-19 and its severity prediction.

In this proposed method, we have used numerous available CNN-based standard pre-trained models on ImageNet to get appropriate features and classical machine learning algorithms such as SVMs to get the final predictive model. We have used several deep feature representations extracted through pre-trained deep learning models and conventinals machine learning algorithms such as SVC, XGBC coupled with rigorous valuation criteria to ultimately achieve a predictive model having better generalization performance. We have used stringent evaluation criteria across combinations of deep feature maps and classical machine learning algorithms to come up with a model that has improved generalization. Properly evaluated improved generalization performance of our proposed method called COVIDX over the existing state-of-the-art methods [9,17] suggests its effective use to diagnose COVID-19 and to predict its severity.

CONCLUSIONS AND FUTURE WORK

In the current study, we have designed and developed a machine learning model called COVIDX for the earliest analysis of the COVID-19 pandemic and to predict its severity using digital chest X-ray scans. The rigorous adopted performance evaluation criteria show that our proposed model can effectively be used for COVID-19 diagnosis and to suggest precautionary measures (such as quarantine and RT-PCR test) to avoid a further surge of the infection. This study also reveals that simple shallow machine learning methods with deep feature maps perform better in comparison to complex deep architectures due to data scarcity. Further, we have shown that deep feature maps performed better in comparison to handcrafted texture-based features used in a previous study. The evaluation of our proposed method in a real setting further signifies its role in COVID-19 diagnosis and taking health intensive care decisions. The key outcomes of this study are bulleted below:

• Our proposed showed robust performance under the stringent evaluation setup even in the existence of significant distinctions in the input CXR scans.

• Our proposed system also predicts the severity of COVID-19 infection along with its diagnosis. This is important while taking health intensive care decisions for COVID-19 infected patients and to avoid unnecessary use of limited intensive care facilities.

• We have deployed COVIDX as an openly accessible webserver and also provided source code. This will help the naïve end-users to easily access and use this system for the diagnosis purpose without going into the implementation details.

In future work, we aim to expand the datasets, evaluation and introduce CT scan image modality for this purpose. At the same time, we will try to further increase the efficiency of our proposed method over the task of severity prediction by using extended datasets. We will also try to incorporate this method in medical devices or extend it to other medical tasks.

MATERIALS AND METHODS

In this section, we discuss the details of our experimental strategies to design COVIDX and its evaluation.

Datasets and their preprocessing

The datasets used in this study have been collected from different publicly open online COVID-chest-X-ray image repositories [31,32]. We used two different datasets: (i) dataset for COVID-19 diagnosis and (ii) dataset for COVID-19 severity prediction. For COVID-19 diagnosis, we have a dataset of 576, 1583, and 4273 digital X-ray images (jpg format) of healthy persons, pneumonia patients, and COVID-19 patients, respectively. Whereas to predict COVID-19 severity, we have used a dataset of 114 and 164 X-ray images of severe and mild COVID-19 infected patients, respectively. All the images in both datasets have been preprocessed including image resizing (313 × 313 pixels), de-noising, and contrast stretching [33].

Proposed method

We have proposed a machine-learning-based approach to diagnose COVID-19 infection and predict its severity from raw digital chest X-ray images. The method we have adopted in this study has been shown in Fig. 3. A similar methodology has already been used in our previous study with CT scans [34]. In the following sections, we will discuss the details of this method.

Feature extraction

In image analysis, an efficient feature extraction technique is significantly important to obtain the most relevant details from the image to reduce dimensionality. If we employ a suitable feature extraction technique, then the extracted features will perform well over the desired task. There are generally two techniques involved in the literature for this purpose: (1) extraction of handcrafted texture features, (2) deep learning-based

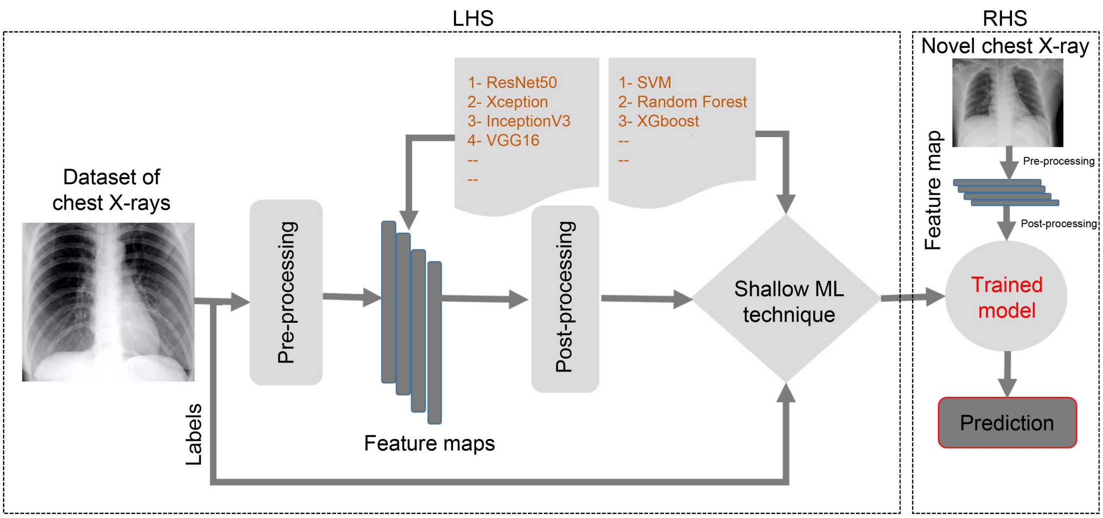

Figure 3. An Adopted framework to develop computer-aided COVID-19 diagnostics and its severity prediction system using machine learning and X-ray scans. LHS (Training: Input raw X-ray image dataset, after preprocessing extract features using pretrained off-the-shelf models, preprocess the features, and then train shallow machine learning algorithms such as SVM); RHS (Testing: Take a novel X-ray image, after extracting its features the trained model will generate the required predictions)

automatically extracted feature maps. It has been proved that deep learning-based feature maps perform better in comparison to handcrafted texture-based features over X-ray images [20]. Therefore, in this study, we only focus on deep learning-based feature maps.

CNN based deep-learning approach, where an automatic feature representation of a digital image can efficiently be learned, is very popular and has efficiently been used previously in various image categorization and investigation tasks. However, to apply deep learning effectively, there is always a prerequisite of a handsome amount of data to efficiently learn the underlying distribution of images in the dataset. This nature of deep learning with data deprivation is quite a challenge to apply it effectively and efficiently in this case where we face the data scarceness. Therefore, to handle the aforementioned issue of data scarceness, we have employed a transfer-learning approach to get a feature representation of CXR scans in our datasets. By using a transfer learning strategy to extract useful feature space from the digital chest X-ray scans in our datasets, we have used various available standard pre-trained on ImageNet CNN-based models. These off-the-shelf pre-trained CNN-based models include DenseNet121 [35], Resnet50 [36], Xception [37], InceptionV3 [38], VGG16 [39], NASNetLarge [40]. We have extracted deep feature maps using these models by simply loading an off-the-shelf model without its fully connected layer by setting the "include_top" argument as "False". We have selected these pre-trained CNN-based models in this study based on their reported generalization capacity. We have applied all the preprocessing and resizing to images in our datasets required by the pre-trained models before employing them to extract the required feature maps.

Problem formulation

We have proposed a technique in this study to diagnose COVID-19 and predict its severity based on machine learning algorithms trained using digital chest X-ray scans. The newness of the offered approach, as deliberated above, is to use a merger of standard pre-trained CNN-based models to get feature maps and traditional machine learning algorithms for the diagnosis and severity prediction tasks. In some way, our proposed framework bears a resemblance to the pattern of transfer learning.

Here, our dataset of digital CXR scans contains instances of the form $(\boldsymbol{X}_i, \boldsymbol{y}_i)$ where $\boldsymbol{X}_i$ is a chest X-ray image and $\boldsymbol{y}_i \in \{+1, -1\}$ is its linked label. For COVID-19 diagnosis, $\boldsymbol{y}_i$ specifies whether $\boldsymbol{X}_i$ signifies COVID-19 (+1) or not (−1) and to predict COVID-19 severity, $\boldsymbol{y}_i$ specifies mild (−1) or severe (+1). We have extracted feature maps denoted by $\boldsymbol{x}_i$ for every chest X-ray image $\boldsymbol{X}_i$ in our datasets. By using these extracted feature maps, we aim to acquire functions that can identify coronavirus infection and elucidate its severity. To learn functions, we have employed three diverse classical machine learning-based algorithms: (i) support vector machine (SVM), (ii) gradient boosting machine (XGBoost), and (iii) random forest (RF) [41–43].

• **Support vector classification (SVC)**

We have employed support vector machines (SVMs) to diagnose COVID-19 and predict its severity by learning a function of the form $f(\boldsymbol{x}) = \langle \boldsymbol{w}, \boldsymbol{x} \rangle$ with $\boldsymbol{w}$ as a weight vector to be learned through training by using the available training instances: $\{(\boldsymbol{x}_i, y_i) | i = 1, 2, \ldots, N\}$. To learn the optimal weight vector $\boldsymbol{w}$ by using SVM, the following optimization problem is required to be solved [42].

$$\min_{w,\xi} \frac{1}{2} \lambda \boldsymbol{w}^2 + \sum_{i=1}^{N} \xi_i$$

$$\text{Such that for all } i : y_i \langle \boldsymbol{w}, \boldsymbol{x}_i \rangle \geqslant 1 - \xi_i,\ \xi_i \geqslant 0 \qquad (1)$$

The objective function is shown in Eq. (1) tries to maximize the margin while curtailing the margin violations (or slacks ξ) [42]. The hyper-parameter $\lambda = \frac{1}{C}$ is used to control the tradeoff between margin violation and margin maximization. We have used both linear and non-linear (radial basis function (RBF)) kernels. Values of λ and γ have been coarsely optimized through grid search with ranges of values via python based machine learning package called Scikit-learn (version: 0.23) [44,45].

• **Random forest classification (RFC)**

Random forest is a supervised machine learning algorithm that builds an ensemble of decision trees, usually trained with the “bagging” method. This technique works by building numerous parallel decision trees while training and produces outputs as a means of the classes of all trees [41]. It performs efficiently on problems with a non-linear relationship in their feature space. While training, RF builds each decision tree by taking a random subset sample of input feature space. RF algorithm has serval hyper-parameters that need to be optimized for better results. We have optimized in this work, through grid search, the maximum number of features required to split a node, the total amount of decision trees, the least number of samples essential to split, and the maximum number of levels in each decision tree [45]. This machine learning technique has been witnessed performing efficiently in many other published studies [34,46–49].

• **XGBoost classification (XGBC)**

XGBoost is a boosting-based ensemble learning technique that chains several weak learners into stronger ones in an iterative way [43,50]. At the core of XGBoost, there is boosting that lessens biases by supervising the model about what errors have been made by previous models and variance by maneuvering multiple models. In the XGBoost technique, each subsequent model is mentored using the residuals (the variance between the predicted and actual values), then models are fitted via subjective differentiable loss function and gradient descent optimization method by pushing the limits of computational resources for efficient throughput. Here, we used trees as default base learners and optimized XGBoost in terms of the number of boosting iterations, the learning rate, booster, maximum depth, and subsample ratio by employing grid search technique and a python-based package called XGBoost 0.7 [34,45,50].

Experimental format

To train a machine learning-based model and to evaluate its performance to diagnose COVID-19 and predict its severity, we have followed the following experimental setup. We have divided the preprocessed images into two sub-sets: train-test set (80%), external validation set (20%), and reported generalization performance using both the sub-sets. We have used a train-test set for 10-fold cross-validation (CV where we split all the images in the subset into 10 groups after shuffling. By using these 10 groups, we have trained 10 models repeatedly with 9 groups as train sets and reported accuracy using each group to be held out as the test set [51]. Similarly, to further confirm the robustness of the generalization performance of our proposed technique, we have used an external validation dataset. To report the predictive performance using an external validation set, we have used the whole train-test set to train the classification models and used the external validation set as the test set. We have computed the area under the precision-recall curve (PR), the area under the ROC curve (ROC), and the F1-measure from the predicted scores and actual labels and reported as accuracy metrics to evaluate the trained model and for its performance assessment [51–53]. We have not reported accuracy as a performance metric because of data imbalance and unevenly correctly classified class importance. While using the 10-fold CV, we have reported the mean value of performance metrics across folds. We have used grid search over training data to determine the optimum values of hyper-parameters of various classification algorithms [45]. We have also evaluated the performance of COVIDX in a real setting using its webserver under the observation of a qualified radiologist.

Statistical analysis

We have also performed the statistical analysis by checking the statistical significance of obtained performance (F1 score) in comparison to the state-of-the-art previously published techniques. For this purpose, we have used Wilcoxon test [54]. The test considers the null hypothesis as the median of the performance scores of the proposed model and previously published techniques are equal. Alternatively, our proposed solution exhibits higher performance. We have used the test statistics at a 95% confidence interval (or α=0.05). We have performed this analysis using an online webserver (STAC Web Platform) [55].

COVIDX webserver

The optimal trained machine learning model in this study has been deployed as an openly accessible webserver. This webserver simply accepts a digital chest X-ray scan to perform coronavirus the required task. After successfully uploading an X-ray image, the users can get COVIDX predictions on the same page. This procedure has been divided into three steps: 1) to determine the uploaded scan reflects a healthy or an unhealthy person (healthy vs unhealthy), 2) if the uploaded image belongs to an unhealthy person then whether it is COVID-19 infection or not (COVID-19 or not), and 3) if it is coronavirus infection then predict its severity (please see Fig. 4). This webserver is freely available at the site of google.

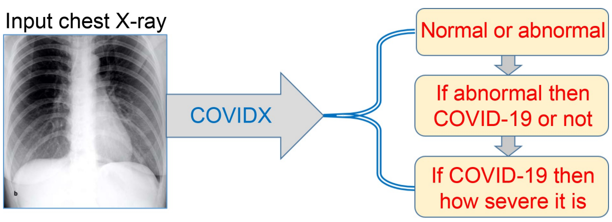

Figure 4. Working flow of COVIDX for the diagnosis of COVID-19 and its severity prediction using chest X-ray scans.

ACKNOWLEDGEMENTS

The authors are highly thankful to all those who have given free access to their compiled and annotated X-ray image datasets.

COMPLIANCE WITH ETHICS GUIDELINES

The authors Wajid Arshad Abbasi, Syed Ali Abbas, Saiqa Andleeb, Maryum Bibi, Fiaz Majeed, Abdul Jaleel and Muhammad Naveed Akhtar declare that they have no conflict of interest or financial conflicts to disclose.

All procedures performed in studies involving animals were in accordance with the ethical standards of the institution or practice at which the studies were conducted, and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

REFERENCES

1. Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., *et al.* (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet, 395, 497−506
2. Kooraki, S., Hosseiny, M., Myers, L. and Gholamrezanezhad, A. (2020) Coronavirus (COVID-19) outbreak: what the department of radiology should know. J. Am. Coll. Radiol., 17, 447−451
3. COVID-19 Map, Johns Hopkins Coronavirus Resource Center. (n.d.). https://coronavirus.jhu.edu/map.html. Accessed: November 27, 2020
4. Coronavirus disease (COVID-19) – World Health Organization, (n.d.). Available from the website of World Health Organization
5. Cheng, S.-C., Chang, Y.-C., Fan Chiang, Y.-L., Chien, Y.-C., Cheng, M., Yang, C.-H., Huang, C.-H. and Hsu, Y.-N. (2020) First case of coronavirus disease 2019 (COVID-19) pneumonia in Taiwan. J. Formos. Med. Assoc., 119, 747−751
6. Commissioner, O., and the Coronavirus Disease 2019 Testing Basics, FDA. (2020). Available from the website of U.S. Food & Drug Administration
7. Sheikhzadeh, E., Eissa, S., Ismail, A. and Zourob, M. (2020) Diagnostic techniques for COVID-19 and new developments. Talanta, 220, 121392
8. Chowdhury, M. E. H., Rahman, T., Khandakar, A., Mazhar, R., Kadir, M. A., Mahbub, Z. B., Islam, K. R., Khan, M. S., Iqbal, A., Al-Emadi, N., *et al.* (2020) Can AI help in screening viral and COVID-19 pneumonia? IEEE Access, 8, 132665–132676
9. Chandra, T. B., Verma, K., Singh, B. K., Jain, D. and Netam, S. S. (2021) Coronavirus disease (COVID-19) detection in chest X-ray images using majority voting based classifier ensemble. Expert Syst. Appl., 165, 113909
10. Chandra, T. B. and Verma, K. (2020) Pneumonia detection on chest X-ray using machine learning paradigm. In: Proceedings of 3rd International Conference on Computer Vision and Image Processing, 21
11. Zhang, N., Wang, L., Deng, X., Liang, R., Su, M., He, C., Hu, L., Su, Y., Ren, J., Yu, F., *et al.* (2020) Recent advances in the detection of respiratory virus infection in humans. J. Med. Virol., 92, 408−417
12. Asnaoui, K. E., Chawki, Y. and Idri, A. (2020) Automated methods for detection and classification pneumonia based on X-ray images using deep learning. ArXiv, 2003.14363
13. Jaiswal, A. K., Tiwari, P., Kumar, S., Gupta, D., Khanna, A. and Rodrigues, J. J. P. C. (2019) Identifying pneumonia in chest X-

rays: A deep learning approach. Measurement, 145, 511−518
14. Pesce, E., Joseph Withey, S., Ypsilantis, P.-P., Bakewell, R., Goh, V. and Montana, G. (2019) Learning to detect chest radiographs containing pulmonary lesions using visual attention networks. Med. Image Anal., 53, 26−38
15. Xue, Z., Jaeger, S., Antani, S., Long, L. R., Karargyris, A., Siegelman, J., Folio, L. R. and Thoma, G. R. (2018) Localizing tuberculosis in chest radiographs with deep learning. In: Proc. SPIE 10579, Medical Imaging 2018: Imaging Informatics for Healthcare, Research, and Applications, 105790U
16. LeCun, Y., Haffner, P., Bottou, L. and Bengio, Y. (1999) Object recognition with gradient-based learning. In: Shape, Contour and Grouping in Computer Vision, Forsyth, D.A., Mundy, J.L., di Gesú, V. and Cipolla, R. (Eds.), pp. 319–345. Springer, Berlin, Heidelberg
17. Wang, D., Mo, J., Zhou, G., Xu, L. and Liu, Y. (2020) An efficient mixture of deep and machine learning models for COVID-19 diagnosis in chest X-ray images. PLoS One, 15, e0242535
18. Khan, A. I., Shah, J. L. and Bhat, M. M. (2020) CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images. Comput. Methods Programs Biomed., 196, 105581
19. Minaee, S., Kafieh, R., Sonka, M., Yazdani, S. and Jamalipour Soufi, G. (2020) Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning. Med. Image Anal., 65, 101794
20. Zare, M. R., Alebiosu, D. O. and Lee, S. L. (2018) Comparison of handcrafted features and deep learning in classification of medical X-ray images. In: 2018 Fourth International Conference on Information Retrieval and Knowledge Management (CAMP), pp. 1–5
21. Abbas, A., Abdelsamea, M. M. and Gaber, M. M. (2020) Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. Appl. Intell., 51, 854−864
22. Afshar, P., Heidarian, S., Naderkhani, F., Oikonomou, A., Plataniotis, K. N. and Mohammadi, A. (2020) COVID-CAPS: A capsule network-based framework for identification of COVID-19 cases from X-ray images. Pattern Recognit. Lett., 138, 638−643
23. Al-Waisy, A. S., Al-Fahdawi, S., Mohammed, M. A., Abdulkareem, K. H., Mostafa, S. A., Maashi, M. S., Arif, M. and Garcia-Zapirain, B. (2020) COVID-CheXNet: hybrid deep learning framework for identifying COVID-19 virus in chest X-rays images. Soft Comput., 1−16
24. Ardakani, A. A., Kanafi, A. R., Acharya, U. R., Khadem, N. and Mohammadi, A. (2020) Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks. Comput. Biol. Med., 121, 103795
25. Jain, R., Gupta, M., Taneja, S. and Hemanth, D. J. (2020) Deep learning based detection and analysis of COVID-19 on chest X-ray images. Appl. Intell., 51, 1690−1700
26. Karakanis, S. and Leontidis, G. (2021) Lightweight deep learning models for detecting COVID-19 from chest X-ray images. Comput. Biol. Med., 130, 104181
27. Murugan, R. and Goel, T. (2021) E-DiCoNet: Extreme learning machine based classifier for diagnosis of COVID-19 using deep convolutional network. J. Ambient Intell. Human Comput., 12, 8887−8898
28. Ozturk, T., Talo, M., Yildirim, E. A., Baloglu, U. B., Yildirim, O. and Rajendra Acharya, U. (2020) Automated detection of COVID-19 cases using deep neural networks with X-ray images. Comput. Biol. Med., 121, 103792
29. Panwar, H., Gupta, P. K., Siddiqui, M. K., Morales-Menendez, R. and Singh, V. (2020) Application of deep learning for fast detection of COVID-19 in X-rays using nCOVnet. Chaos Solit. Frac., 138, 109944
30. Toğaçar, M., Ergen, B. and Cömert, Z. (2020) COVID-19 detection using deep learning models to exploit social mimic optimization and structured chest X-ray images using fuzzy color and stacking approaches. Comput. Biol. Med., 121, 103805
31. Wang, L., Lin, Z. Q. and Wong, A. (2020) COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. Sci. Rep., 10, 19549
32. Cohen, J. P., Morrison, P., Dao, L., Roth, K., Duong, T. Q. and Ghassemi, M. (2020) COVID-19 image data collection: prospective predictions are the future. ArXiv, 2006.11988
33. Chandra, T. B. and Verma, K. (2020) Analysis of quantum noise-reducing filters on chest X-ray images: A review. Measurement, 153, 107426
34. Abbasi, W. A., Abbas, S. A., Andleeb, S., Ul Islam, G., Ajaz, S. A., Arshad, K., Khalil, S., Anjam, A., Ilyas, K., Saleem, M., *et al.* (2021) COVIDC: An expert system to diagnose COVID-19 and predict its severity using chest CT scans: application in radiology. Inform. Med. Unlocked, 23, 100540
35. Huang, G., Liu, Z., van der Maaten, L. and Weinberger, K. Q. (2018) Densely connected convolutional networks. ArXiv, 1608.06993
36. He, K., Zhang, X., Ren, S. and Sun, J. (2015) Deep residual learning for image recognition. ArXiv, 1512.03385
37. Chollet, F. (2017) Xception: Deep learning with depthwise separable convolutions. ArXiv, 1610.02357
38. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z. (2015) Rethinking the inception architecture for computer vision. ArXiv, 1512.00567
39. Simonyan, K. and Zisserman, A. (2015) Very deep convolutional networks for large-scale image recognition. ArXiv, 1409.1556
40. Zoph, B., Vasudevan, V., Shlens, J. and Le, Q. V. (2018) Learning transferable architectures for scalable image recognition. ArXiv, 1707.07012
41. Breiman, L. (2001) Random forests. Mach. Learn., 45, 5−32
42. Cortes, C. and Vapnik, V. (1995) Support-vector networks. Mach. Learn., 20, 273−297
43. Friedman, J. H. (2001) Greedy function approximation: A gradient boosting machine. Ann. Stat., 29, 1189−1232
44. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion,

B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., *et al*. (2011) Scikit-learn: machine learning in python. J. Mach. Learn. Res., 12, 2825−2830
45. Bergstra, J. and Bengio, Y. (2012) Random search for hyper-parameter optimization. J. Mach. Learn. Res., 13, 281−305
46. Abbasi, W. A., Hassan, F. U., Yaseen, A. and Minhas, F. U. A. A. (2020) ISLAND: *In-silico* prediction of proteins binding affinity using sequence descriptors. BioData Min., 13, 20
47. Li, H., Leung, K.-S., Wong, M.-H. and Ballester, P. J. (2014) Substituting random forest for multiple linear regression improves binding affinity prediction of scoring functions: cyscore as a case study. BMC Bioinformatics, 15, 291
48. Ballester, P. J. and Mitchell, J. B. O. (2010) A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking. Bioinformatics, 26, 1169−1175
49. Moal, I. H., Agius, R. and Bates, P. A. (2011) Protein-protein binding affinity prediction on a diverse set of structures. Bioinformatics, 27, 3002−3009
50. Chen, T. and Guestrin, C. (2016) XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 785
51. Abbasi, W. A. and Minhas, F. U. A. A. (2016) Issues in performance evaluation for host-pathogen protein interaction prediction. J. Bioinform. Comput. Biol., 14, 1650011
52. Davis, J. and Goadrich, M. (2006) The relationship between precision-recall and ROC curves. In: Proceedings of the 23rd International Conference on Machine Learning, ACM, 233
53. Tharwat, A., (2018) Classification assessment methods. App. Comput. and Inform., 17, 168–192
54. Wilcoxon, F. (1992) Individual comparisons by ranking methods. In: Breakthroughs in Statistics: Methodology and Distribution. Kotz, S. and Johnson, N. L. (Eds.), pp. 196–202. Springer, New York
55. Rodriguez-Fdez, I., Canosa, A., Mucientes, M. and Bugarin, A. (2015) STAC: A web platform for the comparison of algorithms using statistical tests. In: 2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1–8

