IEEE *Access*

Multidisciplinary : Rapid Review : Open Access Journal

DL-CRC: Deep Learning-based Chest Radiograph Classification for COVID-19 Detection: A Novel Approach

SADMAN SAKIB[1], TAHRAT TAZRIN[1], MOSTAFA M. FOUDA[2,3], (Senior Member, IEEE), ZUBAIR MD. FADLULLAH[1,4], (Senior Member, IEEE), AND MOHSEN GUIZANI[5], (Fellow, IEEE)

[1]Department of Computer Science, Lakehead University, Thunder Bay, Canada
[2]Department of Electrical and Computer Engineering, College of Science and Engineering, Idaho State University, Pocatello, ID, USA
[3]Department of Electrical Engineering, Faculty of Engineering at Shoubra, Benha University, Egypt
[4]Thunder Bay Regional Health Research Institute (TBRHRI), Thunder Bay, Ontario, Canada
[5]Department of Computer Science & Engineering, College of Engineering, Qatar University, Doha, Qatar

Corresponding author: Sadman Sakib (e-mail: ssak2921@lakeheadu.ca).

ABSTRACT With the exponentially growing COVID-19 (coronavirus disease 2019) pandemic, clinicians continue to seek accurate and rapid diagnosis methods in addition to virus and antibody testing modalities. Because radiographs such as X-rays and computed tomography (CT) scans are cost-effective and widely available at public health facilities, hospital emergency rooms (ERs), and even at rural clinics, they could be used for rapid detection of possible COVID-19-induced lung infections. Therefore, toward automating the COVID-19 detection, in this paper, we propose a viable and efficient deep learning-based chest radiograph classification (DL-CRC) framework to distinguish the COVID-19 cases with high accuracy from other abnormal (e.g., pneumonia) and normal cases. A unique dataset is prepared from four publicly available sources containing the posteroanterior (PA) chest view of X-ray data for COVID-19, pneumonia, and normal cases. Our proposed DL-CRC framework leverages a data augmentation of radiograph images (DARI) algorithm for the COVID-19 data, by adaptively employing the generative adversarial network (GAN) and generic data augmentation methods to generate synthetic COVID-19 infected chest X-ray images to train a robust model. The training data consisting of actual and synthetic chest X-ray images are fed into our customized convolutional neural network (CNN) model in DL-CRC, which achieves COVID-19 detection accuracy of 93.94% compared to 54.55% for the scenario without data augmentation (i.e., when only few actual COVID-19 chest X-ray image samples are available in the original dataset). Furthermore, we justify our customized CNN model by extensively comparing it with widely adopted CNN architectures in the literature, namely ResNet, Inception-ResNet v2, and DenseNet that represent depth-based, multi-path-based, and hybrid CNN paradigms. The encouragingly high classification accuracy of our proposal implies that it can efficiently automate COVID-19 detection from radiograph images to provide a fast and reliable evidence of COVID-19 infection in the lung that can complement existing COVID-19 diagnostics modalities.

INDEX TERMS COVID-19, convolutional neural network (CNN), deep learning, generative adversarial network (GAN), pneumonia.

I. INTRODUCTION

THE severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), first observed in Wuhan, China, turned into a global pandemic of COVID-19 (coronavirus disease 2019) [1]. COVID-19 has a destructive impact on the well-being of people, particularly senior citizens and patients with underlying health conditions and compromised immunity levels. By mid-July 2020, the COVID-19 pandemic already contributed to over 570,000 mortalities and more than 13 million cases of COVID-19 infection [2]. A critical step
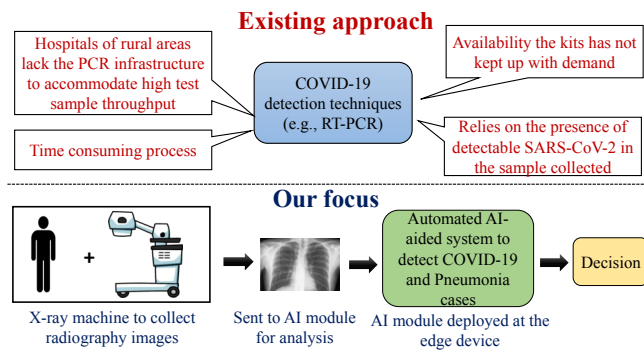
FIGURE 1. Challenges of existing system and our research focus for COVID-19 screening in rural areas.

to combat the pandemic is to effectively detect COVID-19 infected patients as early as possible so that they may receive appropriate attention and treatment. Early detection of COVID-19 is also important to identify which patients should isolate to prevent the community spread of the disease. However, considering the recent spreading trend of the COVID-19, an effective detection remains a challenging task, particularly in communities with limited medical resources. While the reverse transcription polymerase chain reaction (RT-PCR) test-kits emerged as the main technique for COVID-19 diagnosis, chest X-ray (chest X-ray), computed tomography (CT) scans, and biomarkers (i.e. high C-reactive protein (CRP), low procalcitonin (PCT), low lymphocyte counts, elevated Interleukin-6 (IL6), and Interleukin-10 (IL10)) are also being increasingly considered by many nations to aid diagnosis and/or provide evidence of more severe disease progression [3]–[5].

As depicted in Fig. 1, the existing system for detecting COVID-19 using the aforementioned virus and antibody testing modalities is time-consuming and requires additional resources and approval, which can be a luxury in many developing communities. Hence, at many medical centers, the test kits are often unavailable. Due to the shortage of kits and false-negative rate of virus and antibody tests, the authorities in Hubei Province, China momentarily employed radiological scans as a clinical investigation for COVID-19 [**?**]. Motivated by this, several researchers and sources recommend the use of chest radiograph for suspected COVID-19 detection [6]–[8]. Therefore, radiologists can observe COVID-19 infected lung characteristics (e.g., ground glass opacities and consolidation) by harnessing non-invasive techniques such as CT scan or chest X-ray. However, it is difficult to differentiate the COVID-19-inflicted features from those of community acquired bacterial pneumonia [9]. Therefore, for many patients, manual inspection of the radiograph data and accurate decision making can be overwhelming for the radiologists, and an automated classification technique needs to be developed. In addition, radiologists may get infected and need to isolate that may impact rural communities with a limited number of hospitals, radiologists, and caregivers. Moreover, as the second wave of COVID-19 is anticipated

in the fall of 2020, preparedness to combat such scenarios will involve increasing use of portable chest X-ray devices due to widespread availability and reduced infection control issues that currently limit CT utilization [9]. Therefore, as depicted in Fig. 1, in this paper, to automate the COVID-19 detection using X-ray images, we aim to develop an artificial intelligence (AI)-based smart chest radiograph classification framework to distinguish the COVID-19 cases with high accuracy from other abnormal (e.g., pneumonia) and normal cases. In this vein, the main contributions of the paper can be summarized as follows:

- A deep learning-based predictive analytics approach is employed to propose a smart and automated classification framework for predicting COVID-19, pneumonia, and normal cases. Our proposed deep learning-based chest radiograph classification (DL-CRC) framework consists of a data augmentation of radiograph images (DARI) algorithm and a customized convolutional neural network model.

- A uniquely compiled dataset from multiple publicly available sources is prepared with radiographs of healthy (normal), COVID-19, and pneumonia cases reported to date. The limited number of COVID-19 instances in the dataset is identified as the prime reason for training bottleneck of deep learning algorithms. As a solution, our proposed DARI algorithm essentially combines a customized generative adversarial network (GAN) model with several generic augmentation techniques to generate synthetic radiograph data to overcome the COVID-19 class imbalance problem due to limited dataset availability.

- We train a customized CNN model based on combined real and synthetic radiograph images that contributes to significantly improved accuracy of 93.94% in contrast with 54.55% when only actual COVID-19 instances in public datasets are used for training. While chest X-ray is regarded as a less sensitive modality in detecting COVID-19 infection in lungs compared to CT scans in the literature [9], we demonstrate the good performance of our custom CNN model in identifying COVID-19 cases in the real dataset with high accuracy implying that our approach nullifies the need for using expensive CT scan machines because the COVID-19 detection accuracy using our custom CNN model is much higher compared to the reported baseline [9].

- We rigorously analyze the computational complexity of the DARI, training, and running/inference steps of our proposed DL-CRC framework. The analyses, further corroborated by experimental results, reveal that our proposed methodology leads to significantly lower training time, and particularly much improved inference time, which is crucial for deploying the trained model into portable X-ray devices for fast and reliable COVID-19 feature detection in lung radiographs.

- The performance of our customized CNN model is ex-

tensively compared with the state-of-the-art CNN architectures in the literature (i.e., depth-based CNNs, multipath-based CNNs, and so forth) [10]. Our proposal is demonstrated to substantially outperform the contemporary models in terms of classification efficiency.

The remainder of the paper is organized as follows. Section II surveys the relevant research work regarding COVID-19 and the relevant use of AI. The problem of traditional COVID-19 detection and challenges associated with it to apply in developing communities is discussed in section III. Our proposed input representation and deep learning model are presented in section IV. The performance of our proposal is evaluated in section V and extensively compared with those of well-known CNN architectures. Some of the limitations of the study is briefly explored in section VI. Finally, section VII concludes the paper.

II. RELATED WORK

This section explores the relevant research work in the literature from two perspectives, i.e., imaging modalities for COVID-19 detection, and AI-based analysis of radiograph samples.

A. IMAGING MODALITIES FOR COVID-19 DETECTION

Most nations had to take measures to react to the sudden and rapid outbreak of COVID-19 within a relatively short period of time. According to [11], radiology departments have started to focus more on preparedness rather than diagnostic capability, after sufficient knowledge was gathered regarding COVID-19. The study in [5] stated the resemblance of COVID-19 with other diseases caused by other coronavirus variants such as the severe acute respiratory syndrome (SARS) and the middle east respiratory syndrome (MERS). The importance of a tracking the lung condition of a recovering coronavirus patient using CT scans was also mentioned in the study. Chest imaging techniques were highlighted to be a crucial technique for detecting COVID-19 by capturing the bilateral nodular and peripheral ground glass opacities in the lung radiograph images [12].

B. AI-BASED RADIOGRAPH ANALYSIS

The application of AI, for early detection, diagnosis, monitoring, and developing vaccines for COVID-19, were elaborately discussed in [13]. Several research work exist in the literature that exploited various deep learning techniques on X-ray data to demonstrate reasonable performance [14]–[17]. In [18], a model, referred to as DarkCovidNet, for early detection of COVID-19 was proposed which utilized 17 convolutional layers to perform binary and multi-class classification involving normal, COVID, and pneumonia cases. While the model reported an overall accuracy of 98.08% for the binary classification and 87.02% for multi-class classification, our reconstruction of the DarkCovidNet using multiple datasets indicated overtraining and much lower accuracy when non-biased test data are presented to the model. Several other papers applied deep learning models on CT scan images to detect and monitor COVID-19 features in the radiograph data [19], [20]. Ardakani et. al in [21] employed implemented the state-of-the-art CNN architectures such as AlexNet, ResNet-18, ResNet-50, ResNet-101, SqueezeNet, VGG-16, VGG-19, MobileNet-V2, GoogleNet, and XceptionCT to differentiate between COVID-19 and non-COVID-19 cases. Their experiments showed that deep learning could be considered as a feasible technique for identifying COVID-19 from radiograph images. To avoid poor generalization and overfitting due to lack of COVID-19 samples in available datasets, a GAN model was used in [22] to generate synthetic data, which achieved a dice coefficient of 0.837. The applicability of GAN for COVID-19 radiograph data synthesis can be confirmed from the broader spectrum of GAN applications on various medical data according to the survey in [23]. The survey identified various unique properties of GAN such as domain adaptation, data augmentation, and image-to-image translation that encouraged researchers to adopt it for image reconstruction, segmentation, detection, classification, and cross-modality synthesis for various medical applications.

III. PROBLEM STATEMENT

With the rapidly surging pandemic, the demand for efficient COVID-19 detection has dramatically increased. The lack of availability of COVID-19 viral and antibody test-kits, and the time required to obtain the test results (in the order of days to weeks) in many countries are posing a great challenge in developing/rural areas with less equipped hospitals or clinics. For instance, in many developing countries, hospitals do not have sufficient COVID-19 test-kits, and therefore, they require the assistance of more advanced medical centers to collect, transport, and test the samples. This creates a bottleneck in mass testing for COVID-19. Therefore, to meet the daily demand for an enormous amount of new test cases, an automated and reliable complementary COVID-19 detection modality is necessary, particularly to confront the second wave of the pandemic. Radiograph image utilization for initial COVID-19 screening may play a pivotal role in areas with inadequate access to a viral/antibody testing. In several studies, CT scans were used for analyzing and detecting features of COVID-19 [24] due to higher resolution of features of ground glass opacities and lung consolidation compared to chest X-ray images. However, due to infection control matters associated with patient transport to CT suites, relatively high cost (for procurement, operation and maintenance of CT equipment), and the limited number of CT machines in developing/rural areas, CT scan is not a practical solution for detecting COVID-19 [9]. On the other hand, chest X-ray can be employed to identify COVID-19 or other pneumonia cases as a more practical and cost-effective solution because X-ray imaging equipment are pervasive at hospital ERs, public healthcare facilities, and even rural clinics. Even for trained radiologists, detecting chest X-ray images pose challenges to distinguish between features of COVID-19 and community acquired bacterial pneumonia [9]. Moreover, the influx of patients into hospital ERs during pandemic, manual inspection
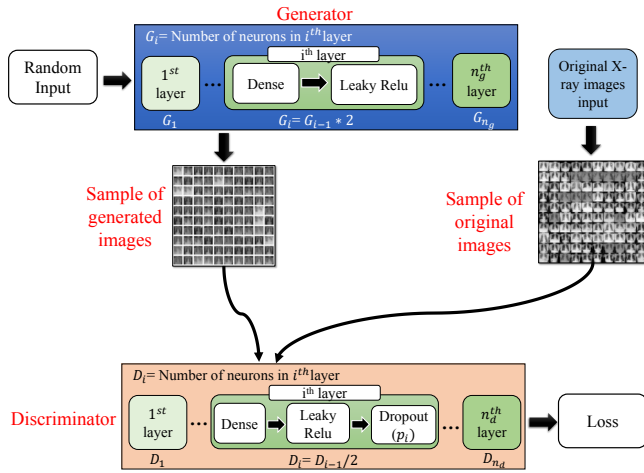
FIGURE 2. Our customized generative adversarial network (GAN) model for data augmentation.

of radiograph data and accurate decision making can lead to a formidable tradeoff between detection time and accuracy that can overwhelm the radiologist department. Therefore, an automated classification technique needs to be designed. As the second wave of COVID-19 is expected in many countries, preparedness to combat the pandemic will involve increasing use of portable chest X-ray devices due to widespread availability and reduced infection control issues that currently limit CT utilization [9]. In the following section, we address the aforementioned problem and present a deep learning-based approach to effectively solve the problem.

IV. PROPOSED DEEP LEARNING-BASED CHEST RADIOGRAPH CLASSIFICATION (DL-CRC) FRAMEWORK

Deep learning in smart health analytics is a prominent interdisciplinary field that merges computer science, biomedical engineering, health sciences, and bioinformatics. Various medical imaging devices have a dedicated image and signal analysis and processing module, on which deep learning-based models can be implemented to provide accurate, real-time inferences. Motivated by this, we conceptualize a deep learning-based chest radiograph classification (DL-CRC) framework, which can used for automating COVID-19 detection from radiograph images.

Our proposed DL-CRC framework consists of two components: (*i*) the data augmentation of radiology images (DARI) algorithm, and (*ii*) a deep learning model. Our proposed DARI algorithm generates synthetic X-ray images by adaptively switching between a customized GAN architecture and generic data augmentation techniques such as zoom and rotation. The synthetic X-ray images are combined with the actual radiograph data to build a robust dataset for efficiently training the deep learning model, i.e., the second component of our DL-CRC framework. A custom CNN architecture is designed to construct the deep learning model to carry out automated feature extraction and classification of the

radiograph images.

Next, the details of the proposed DARI algorithm and custom CNN model of our envisioned DL-CRC framework are presented, followed by a rigorous complexity analysis of the proposed methodology in training and inference phases.

A. PROPOSED DARI ALGORITHM

Here, we propose an adaptive data augmentation of radiograph images algorithm, referred to as DARI. Our proposed DARI algorithm performs an on-demand generation of synthetic X-ray images, triggered by class imbalance in the original dataset. The generated synthetic images are combined with actual radiograph images to construct a robust training dataset. This is essential, in the COVID-19 context, where enough representative samples of COVID-19 chest X-ray images are not sufficient in the currently available datasets. DARI leverages a custom GAN model, as depicted in Fig. 2, along with generic data augmentation techniques such as zoom and rotation. The GAN model is invoked if the number of samples in a class is less than a certain predefined threshold (δ). In the GAN model, a generator (G) and a discriminator (D) are trained simultaneously until the discriminator is unable to separate the generated data samples from the original ones. The generator receives random noise as input and produces chest X-ray images, which are, in turn, received by the discriminator. Thus, the GAN can be regarded as a two-player minimax game between a discriminative model (D) and a generative model (G) [25]. By exerting a noisy sample n_x with the data distribution of $p(n_x)$ as the input, the generative network G outputs new data X', distribution of which, denoted by $p(X')$, is supposed to be identical to that of the distribution of original data, $p(X)$. The discriminative network, D, is employed to distinguish the true data sample X with the distribution of $p(X)$ and the generated sample X' with a distribution of $p(X')$. Then, this adversarial training process can be formulated as follows,

$$\min_G \max_D V(D, G) = \mathbb{E}_{X \sim p(X)} \log(D(X)) \\ + \mathbb{E}_{n_x \sim p(n_x)} \log(1 - D(n_x)). \quad (1)$$

We customize the GAN model for chest X-ray image augmentation as follows. The generator is constructed with a stack of n_g hidden layers. Each layer comprises a dense layer, followed by Leaky Rectified Linear Unit (LeakyReLU) as the activation function. In each successive layer (i^{th}) of the generator, the number of neuron units (i.e., nodes) is twice the number of nodes in the preceding layer. On the other hand, in the discriminator model, it receives collections of original (X) and generated (X') X-ray radiograph data with COVID-19 infected lung images. Here, the inputs to the discriminator are $X = [x_1, x_2, \ldots x_n]$ and $X' = [x'_1, x'_2, \ldots x'_n]$, where each x_i represents an original image while each x'_i denotes an augmented chest X-ray image. Similar to the generator, the discriminator's structure also consists of n_d hidden layers, and each ith layer contains a sequence of a dense layer with LeakyReLU as the activation function [26]. A dropout

layer is then included. Let p_i denote the dropout rate. The number of nodes in each ith layer is denoted by D_i. Note that $D_i = \frac{1}{2} \cdot D_{i-1}$. The discriminator aims to optimize the loss function by distinguishing generated images from the original ones. Our custom GAN model is trained for ξ_{max} number of iterations, where $\xi_{max} \in \mathbb{Z}^+$. The detailed steps of our proposed DARI algorithm are presented in Algorithm 1. Here, we either invoke the GAN or a more generic type of data augmentation, based upon a given condition as illustrated in Algorithm 1. This procedure takes two inputs: (i) type of augmentation, and (ii) data for augmentation. For one condition, the proposed GAN model gets executed from steps 2 to 22. When the other condition is fulfilled, the generic data augmentation is performed as described in steps 23 to 25, which includes enlarging the image by Z quantity and rotating by θ amount.

B. PROPOSED CUSTOM CNN MODEL FOR COVID-19 DETECTION IN X-RAY IMAGES

Next, we need to train a deep learning model which can take advantage of the robust dataset obtained from our proposed DARI algorithm in section IV-A. Since the problem can be regarded as a classification task of normal, COVID-19, and other abnormal cases (e.g., pneumonia), we investigate the contemporary deep learning architectures suited for classification. In contrast with other variants of deep learning architectures (i.e., long-short term memory (LSTM), deep belief networks, and so forth) and extreme learning machines, CNNs are regarded as the most powerful deep learning architecture for image classification. Therefore, we explore the robust CNN models recently employed to gain reasonable classification accuracy with chest X-ray data [18]. By applying the contemporary CNN models on the latest dataset compiled from four public repositories, we realize that their reported performances are constrained by overfitting and influenced by biased test data. To address this issue, we propose a two-dimensional (2-D), custom CNN model for classifying X-ray images to predict COVID-19 cases as depicted in Fig. 3. The 2-D CNN structure is utilized to learn the discriminating patterns automatically from the radiograph images.

The proposed CNN model consists of three components. The first component is a stack of n_c convolution layers while the second segment consists of n_d fully connected layers. The final component is responsible for generating the output probability. At first, the convolution layers (i.e., the first component of the model) receive radiograph images (X) as input, identify discriminative features from the input examples, and pass them to the next component for the classification task. Each ith layer among the n_c convolution layers consists of a filter size of z^i. Initially, the filter size is set to x_{i_r} in the 1st layer, and it is decreased by λ in each successive layer. In the forward pass, the convolution operation is performed between the input image and the filter coefficients using Eq. 2. Here, x_{ij}^l and w_{ij}^l denote the output and the filter weights of the lth layer, respectively.
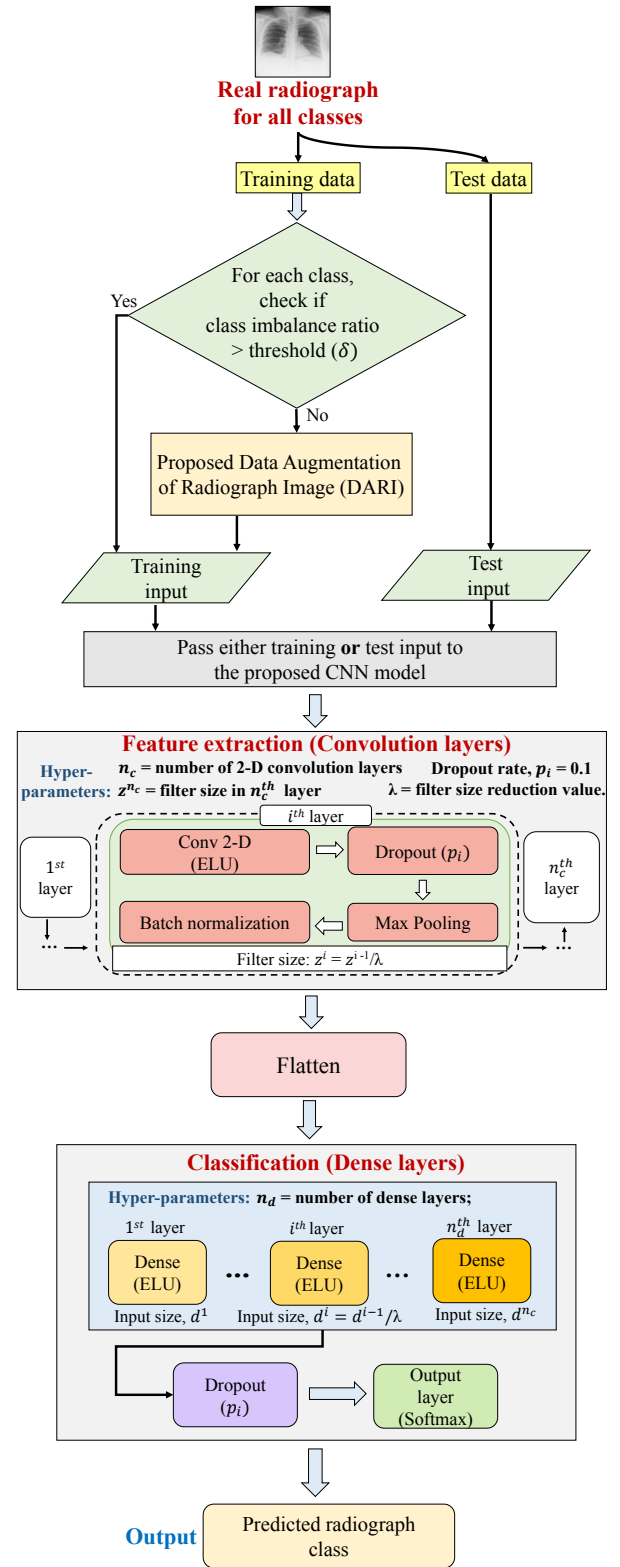

FIGURE 3. Proposed DL-CRC framework consisting of our envisioned DARI algorithm and custom CNN model. (1) The test data is obtained by splitting the original images that are not used for training. (2) DARI algorithm adaptively uses GAN and generic data augmentation techniques to generate synthetic chest X-ray images which are combined with the remaining original radiograph images to construct a robust training dataset. (3) The training input is passed to our customized CNN model, which performs automated feature extraction and classification.

Algorithm 1: Data augmentation of radiograph images (DARI)

```
 Input:  type (type of data augmentation,
         possible values 'generic',
         'GAN'), D (collection of data
         for augmentation)
 Output: γ (augmented sample data)
 1  γ ← ∅
 2  if (type='GAN') then
 3  │   Initialize ξ_max (maximum number of
    │     epochs), B (mini-batch size), and
    │     n_aug (number of data to augment)
 4  │   m_g ← construct generator model as
    │     depicted in Fig. 2
 5  │   m_d ← construct discriminator model
    │     as depicted in Fig. 2
 6  │   foreach e ∈ ξ_max do
 7  │   │   for (i=1 to B) do
 8  │   │   │   n_x ← generate n_aug samples of
    │   │   │     random noise to initialize the
    │   │   │     generator
 9  │   │   │   g_i ← generate image by passing
    │   │   │     n_x to the generator m_g
10  │   │   │   r_i ← select random set of
    │   │   │     samples from D
11  │   │   │   X* ← construct collection from
    │   │   │     generated (g_i) and original
    │   │   │     samples (r_i)
12  │   │   │   m_d ← update the discriminator
    │   │   │     model by batch training using
    │   │   │     X*
13  │   │   end for
14  │   │   n_x ← generate n_aug samples of
    │   │     random noise
15  │   │   m_g ← update the generator model
    │   │     parameters
16  │   │   if e=ξ_max then
17  │   │   │   γ ← generate collection of
    │   │   │     augmented images by using n_x
18  │   │   │   foreach img ∈ γ do
19  │   │   │   │   save img to corresponding
    │   │   │   │     directory
20  │   │   │   end foreach
21  │   │   end if
22  │   end foreach
23  else
24  │   γ ← augment data by applying zooming
    │     rate of Z and rotation of θ on each
    │     item from data collection D
25  end if
26  return γ
```

$$x_{ij}^l = \sum_{i \in x_{i_r}, j \in x_{i_c}} \Omega(x_{ij}^{l-1} \times w_{ij}^l). \qquad (2)$$

Hyper-parameter tuning is conducted to select the optimal activation function, Ω, as shown in in Eq. 2. The activation function considers a constant, denoted by $\alpha > 0$.

Next, we apply a dropout of rate p_i as the regularization technique that will assist the network in evading overfitting and achieve better model generalization by randomly disre-

garding randomly selected neurons in the hidden layers [27]. To reduce the feature size and computational power need, we introduce the max-pooling layer with a pool size of $k_i = (k_r^i, k_c^i)$ in the hidden layers where k_i is set to a fraction μ of the initial dimension of the input x_i. The max-pooling layers assist the model in capturing abstract spatial information more robustly and enhancing the model's generalization ability of the model [28]. The output features of the convolution layers are converted into a one-dimensional (1-D) vector by flattening the layer, and then forwarded to the stack of n_d fully-connected or dense layers for the automated classification stage. The number of nodes in the first dense layer is equal to x_{i_r}, and it is decreased by a factor of λ in each successive i^{th} layer with respect to the number of nodes in the previous layer. The output of the nth dense layer is propagated through a dropout layer of rate p_i.

Finally, the output layer computes the probability of the input x_i belonging to each class. The learning is set to a constant η_c throughout the training of the model.

The classification task receives radiograph samples as input $X = [x_1, x_2, \ldots x_n]$, and outputs a sequence of labels $Y = [y_1, y_2, \ldots y_n]$. Here, each x_i corresponds to the pixel values of the input images. On the other hand, each y_i denotes a distinct class. Each x_i has the dimension of $(x_{i_r}, x_{i_c}, \vartheta_i)$. In this case, x_{i_r}, x_{i_c}, and ϑ_i denote the image height, width, and the number of channels for the ith sample.

The augmented and real samples are passed to the training data during the training phase, and some part of the real samples are considered as the test dataset during the testing phase.

C. TRAINING AND RUNNING PHASES OF PROPOSED DL-CRC

From hereon, we discuss the steps of the training and running phases of our proposed DL-CRC algorithm.

The steps of the training phase of our proposed DL-CRC framework is presented in Algorithm 2. The training stage of DL-CRC commences from Algorithm 2, which takes C, k, B, λ, and δ as inputs to our custom CNN model. The description of each input parameter is provided in the input section of the algorithm. Steps 1 to 3 of Algorithm 2 initialize the required parameters. In steps 4 to 10, all data are loaded from location, and the test data are split by the ratio of λ to be utilized in the running phase for evaluating the model. Initially, all data are stored in the training directory. Hence, they are loaded from the location of training data. Steps 11 to 20 are responsible for checking whether any data augmentation is required or not, and accordingly preparing all the training and validation data from the dataset. Specifically, steps 13 to 15 check whether the training data in any class is less than a predefined threshold δ or not, based on the condition if it can exploit the proposed data augmentation of radiograph images (DARI) algorithm described in Algorithm 1. Our customized CNN model is trained in steps 21-27, utilizing the model structure illustrated in Fig. 3. At the penultimate step, the trained model (M_t) is stored for further testing

---

Algorithm 2: Training phase (DL-CRC)

Input: C (collection for training, testing, and validation data location), k (number of fold in cross-validation), ξ (number of epoch), B (mini-batch size), λ (test ratio), δ (threshold value for class imbalance ratio), $\mathcal{N}$ (total number of samples across all classes)

Output: M_t (Trained model)

```
 1  M_t ← ∅
 2  X ← []
 3  Y ← []
 4  X* ← read all data from C[train]
 5  if (len(X*) > 0) then
 6      I* ← generate random values in
            range[0, λ × len(X*)]
 7      foreach index i ∈ I* do
 8          move C[train] + X*[i] to C[test] + X*[i]
 9      end foreach
10  end if
11  foreach class c_i ∈ C[train] do
12      x*_i ← read all data from c_i
13      if (len(x*_i)/N < δ) then
14          x*_i += DARI('gan', x*_i)
15      end if
16      foreach class data ∈ x*_i do
17          X += data
18          Y += c_i
19      end foreach
20  end foreach
21  for (fold no. j=1 to k) do
22      X_train, y_train, X_val, y_val ← set data and
            labels of j^th fold from X, Y
23      X_train += DARI('generic', X_train)
24      X_val += DARI('generic', X_val)
25      M_t ← update the CNN model depicted
            in Fig. 3 by training it using X_train
            for ξ and B
26      evaluate M_t by using X_val, y_val
27  end for
28  save the model parameters of M_t
29  return M_t
```

and validation. Finally, in step 29, the algorithm returns the trained model.

Next, in the running phase, the CNN model of our proposed DL-CRC framework follows Algorithm 3. It receives the location of sample data for inference and returns the predicted class labels (y_{pred}) for the corresponding data. After reading the data from step 1, the pre-trained model (M_t) is loaded in the following step. In step 4, the model M_t is employed to predict the probabilities for a sample test data to be in each of the possible classes. Finally, in the last step, the class with the maximum probability is identified for each sample data, and then returned as a collection of predictions for all the data.

Algorithm 3: Running phase (DL-CRC)

Input: testPath (location of test images)

Output: y_{pred} (prediction of testing samples)

```
1  X_test ← read all data from testPath
2  M_t ← load the saved pre-trained model
3  y_prob ← predict the probabilities of each
       data from X_test
4  y_pred ← argmax(y_prob)
5  return y_pred
```

D. COMPUTATION OVERHEAD ANALYSIS

In the remainder of the section, we rigorously analyze the computational overhead of our proposed model in terms of time-complexity. The analyses are divided into training and running phases.

1) Training Phase

The training phase includes both our proposed DARI (Algorithm 1) for data augmentation and training our customized CNN model (Algorithm 2). Particularly for the analysis of Algorithm 2, we consider that the appropriate hyperparameters of our CNN model are already selected after hyperparameter tuning. We partition the analysis of the training phase into three main segments, i.e., DP (required data preparation), DA (data augmentation), and CNN (the execution of the CNN model). Therefore, the total computational complexity can be expressed as follows.

$$C(T) = O(\text{DP}) + O(\text{DA}) + O(\text{CNN}). \qquad (3)$$

In the first three steps (1-3) of Algorithm 2, where initialization is conducted, the time complexity can be denoted as constant time, $O(1)$. In the 4th step, all the data from the train path are read. So, if there are f_n number of data available to train, the time complexity will be $O(f_n)$. Steps 5-9 split the test data by the λ ratio. Therefore, the complexity associated with these steps is $O(\lambda)$. Hence, the computational complexity of the data preparation phase can be denoted as:

$$O(\text{DP}) = O(3) + O(f_n) + O(\lambda) \approx O(f_n) + O(\lambda). \quad (4)$$

The data augmentation part of the complexity analysis mainly consists of our proposed DARI (Algorithm 1), invoked in steps 13-15 of Algorithm 2. This requires loading data from each class in step 12 that results in the computational complexity of $O(c_l \times f_n^i)$. Here, c_l denotes the number of classes while f_n^i refers to the number of data read from i^{th} class. Then, through steps 13-15, the DARI algorithm is invoked and its complexity is denoted as ODARI. Suppose that n_g and n_d denote the numbers of layers in the generator and discriminator, respectively. Then, the computations required by the generator and the discriminator models can be denoted as G_c (Eq. 5) and D_c (Eq. 6), respectively:

$$G_c = 2(\sum_{i=1}^{n_g} x_g^i \times w_g^i + b_g^i), \qquad (5)$$

$$D_c = 2(\sum_{i=1}^{n_d} x_d^i \times w_d^i + b_d^i). \tag{6}$$

Combining the previous two expressions of G_c and D_c, the overall overhead of DARI (Algorithm 1) is evaluated as follows.

$$O(\text{DARI}) = O(c_l \times \xi_{\max} \times B \times (G_c + D_c)) + O(c_l \times n_{\text{aug}}), \tag{7}$$

where n_{aug}, $\xi_{\max}$, and B denote the number of data to augment, maximum number of epochs, and mini-batch size, respectively.

In steps 16-19 of the training algorithm, assuming the length of each x_i^* as lx_i^*, the computational overhead is $O(\text{lx}_i^*)$. Therefore, the overall complexity of the data augmentation stage can be expressed as:

$$O(\text{DA}) = O(c_l \times f_n^i) + O(\text{DARI}) + O(\text{lx}_i^*). \tag{8}$$

From steps 21 to 27, the training algorithm invokes the adopted 2-D CNN structure. The computational overhead for this part can be derived from Eq. 9:

$$O(\text{CNN}) = O(\text{CNN}_{cl}) + O(\text{CNN}_{dl}), \tag{9}$$

where $O(\text{CNN}_{cl})$ and $O(\text{CNN}_{dl})$ denote the computational overheads in the convolutional layers and dense layers, respectively. If we consider for a layer i, the number of filters in the ith layer z^i, input image x^i with the dimension of (x_r^i, x_c^i) and kernel k^i with the dimension of (k_r^i, k_c^i), then the computational complexity of the convolutional layers can be expressed as:

$$O(\text{CNN}_{cl}) = O(z^i \times (\sum_{i=1}^{n_c} (x_r^i \times x_c^i \times k_r^i \times k_c^i))). \tag{10}$$

After the convolutional layers, for n layers, assuming w^i and b^i are the weight vector and the bias of ith layer, the complexity of the fully connected layers is given by:

$$O(\text{CNN}_{dl}) = O(\sum_{i=1}^{n_d} (x_r^i \times x_c^i \times w^i + b^i)). \tag{11}$$

Hence, combining the aforementioned equations, to finalize the computational complexity of the proposed CNN, we can re-write Eq. 9 as follows:

$$O(\text{CNN}) = O(z^i \times (\sum_{i=1}^{n_c} (x_r^i \times x_c^i \times k_r^i \times k_c^i))) \\ + O(\sum_{i=1}^{n_d} (x_r^i \times x_c^i \times w^i + b^i)). \tag{12}$$

Finally, to determine the total time complexity of the training phase of the DL-CRC algorithm, we can substitute the corresponding values from Eqs. 4, 8, and 12 into Eq. 3.

2) Running Phase

The running phase is conducted to infer classes of each test data using the pre-trained model and then evaluate the model. As shown in Algorithm 3, if we consider the number of test data to be n_{test}, the computational overhead in the testing phase can be given by:

$$C(R) = O(n_{\text{test}}). \tag{13}$$

Eq. 13 demonstrates that the model is able to produce results in linear time. This implies that our proposed DL-CRC framework comprising DARI algorithm and the customized CNN model can be deployed on clinical-grade X-ray machines with image processing capability, computing resources having access to digitized radiograph images from analog X-ray machines, and even portable X-ray machines in movable booths and trucks with adequate shielding and power supply. Thus, our model is viable for automating the radiograph image classification with fast turn-around time for COVID-19 detection.

V. PERFORMANCE EVALUATION

To evaluate the performance of our proposed DL-CRC framework, in this section, we describe the collected datasets used to train our customized CNN model, followed by extensive experimental results and discussion.

A. DATASET PREPARATION

The dataset employed for the supervised radiograph image classification using our proposed DL-CRC framework consists of three classes: COVID-19, pneumonia, and normal chest X-ray images. We collected the dataset using four different existing datasets of Posteroanterior (PA) chest X-rays, and combined those into a single dataset to utilize it for the classification purpose. We developed the dataset from GitHub for COVID-19 X-rays [29], X-ray data collected in this study for cases of pneumonia, and normal images [30], CheXpert dataset collected by Stanford ML group [31], and the rest of the normal and pneumonia chest X-ray images were collected from the dataset in [32]. Table 1 lists the initial class distribution of the collected chest X-ray dataset. The number of samples collected for COVID-19 is significantly lower than the other two classes because this is a novel disease, and at this moment, data regarding COVID-19 is challenging to obtain. In other words, the number of COVID-19 class samples in the merged dataset is lower than the threshold value for class imbalance ratio, δ. Therefore, to overcome the effect of the low amount of COVID-19 data, we employed our proposed DARI algorithm to increase the number of samples. We then applied our proposal along with contemporary CNN models to verify which one yields the best COVID-19 detection performance.

B. PERFORMANCE INDICATORS

To evaluate the classification results, we primarily adopted the combination of three measurement indicators, accuracy,

weighted precision, and weighted F1 score. The accuracy of a test is its ability to correctly differentiate the three cases. Assume that C denotes the number of classes in the considered classification task, $|y_i|$ refers to the number of samples in the ith class, and $|Y|$ indicates the total number of samples in all the classes. Then, the accuracy can be represented as follows.

$$\text{Accuracy} = \frac{\sum_{i=1}^{C}(\text{TP}_i)}{|Y|}. \quad (14)$$

Next, we define the weighted precision. Our aim is to measure how precise the model is in terms of the number of samples actually present in the ith class out of those predicted to be in that class. This number is multiplied by the weight of the ith class to obtain the weight precision as follows.

$$\text{Weighted precision} = \sum_{i=1}^{C}(\frac{|y_i|}{|Y|} \times \frac{\text{TP}_i}{\text{TP}_i + \text{FP}_i}). \quad (15)$$

Next, the weighted F1 score is defined as the weighted average of precision and recall. Although we did not use recall directly as a performance measure, because of using the F1 score, it is implicitly used. The weighted F1 score can be obtained as follows,

$$\text{Weighted F1 score} = \sum_{i=1}^{C}(\frac{|y_i|}{|Y|} \times 2\frac{P_i \times R_i}{P_i + R_i}). \quad (16)$$

Here, P_i and R_i are the precision and recall of ith class, respectively. P_i can be expressed as $\text{TP}_i/(\text{TP}_i + \text{FP}_i)$ and P_i can be denoted as $\text{TP}_i/(\text{TP}_i + \text{FN}_i)$. TP_i, FP_i, and FN_i denotes True Positive, False Positive, and False Negative for i^{th} class respectively. TP_i indicates the number of cases correctly identified to be in the ith class; FP_i represents the number of cases incorrectly identified to be in the ith class, and FN_i denotes the number of cases incorrectly identified as a class other than the ith class. In addition, for evaluating our results more comprehensively we also employed class specific classification accuracy (i.e., normal, COVID-19, and pneumonia detection accuracy) for all three classes.

C. RESULTS AND DISCUSSION

We have followed a systematic approach by applying different techniques to find the optimal model for the classification task. All the experiments were conducted on a workstation with Intel Core i7, 3.00GHz CPU, 16 GB RAM, powered by Nvidia RTX 2060 Graphics Processing Unit (GPU). The simulations were implemented employing Python's Keras and
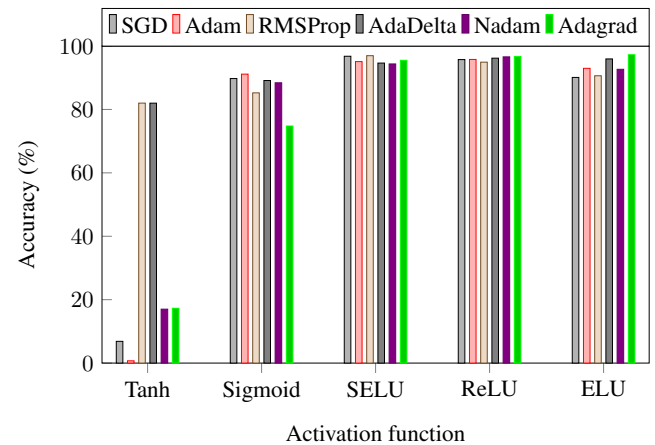

FIGURE 4. Performance in terms of accuracy for different combinations of activation functions and optimizers.

TensorFlow library. The visualization of the experimental results was achieved by utilizing Python's Matplotlib library. During the simulations, we have resized the image samples by setting both x_{i_r} and x_{i_c} to 100 to keep the images consistent in terms of size. The number of channels of the samples (ϑ_i) was set to 1 as the input images were grayscale in nature. The values of x_{i_r} and x_{i_c} were selected based on manual tuning. Using our proposed DARI algorithm, on-demand data augmentation is performed by adaptively employing GAN, rotation (θ) of 5 degrees, and zooming (Z) rate of 0.50. The value of δ was set to 0.1. We systematically constructed three experimental scenarios to conduct a comprehensive performance comparison of our proposed DL-CRC framework consisting of DARI algorithm and our customized CNN models with the state-of-the-art CNN models which have been recently reported to provide reasonable accuracies for COVID-19 detection. The three scenarios, constructed in an incremental fashion, are described below.

1) In our first scenario, we designed our customized deep CNN model architecture depicted in Fig. 3. The parameters of the model were selected based on the results of the grid search technique.

2) In the second scenario, we implemented the proposed DARI algorithm to analyze the effect of the generic and GAN-based data augmentation to train the CNN-based model in a robust fashion to significantly improve the COVID-19 detection accuracy.

3) In the third and final scenario, we trained several state-of-the-art CNN models using different deep learning paradigms on our compiled dataset. The same test data (unknown chest X-ray original images with normal, COVID-19, and pneumonia cases) were presented to the customized CNN model of our proposed DL-CRC framework as well as the contemporary CNN models. The results were used to compare the performances of our proposal and these contemporary models in terms of COVID-19 and pneumonia detection efficiency.

In the first scenario, we implemented the customized CNN

TABLE 1. Brief description of the used dataset for X-ray image classification.

Class label	Number of samples	Modality	View
Normal	27,228		
Pneumonia	5,794	X-ray	Posteroanterior (PA)
COVID-19	209		

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI
10.1109/ACCESS.2020.3025010, IEEE Access

IEEE *Access*

Sakib *et al.*: DL-CRC: Deep Learning-based Chest Radiograph Classification for COVID-19 Detection: A Novel Approach
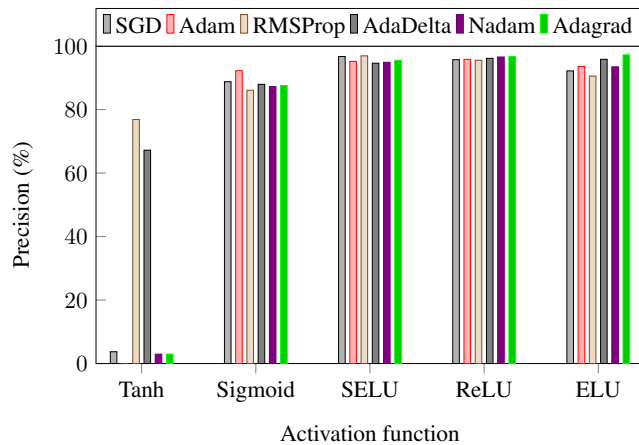
FIGURE 5. Performance in terms of precision for different combinations of activation functions and optimizers.

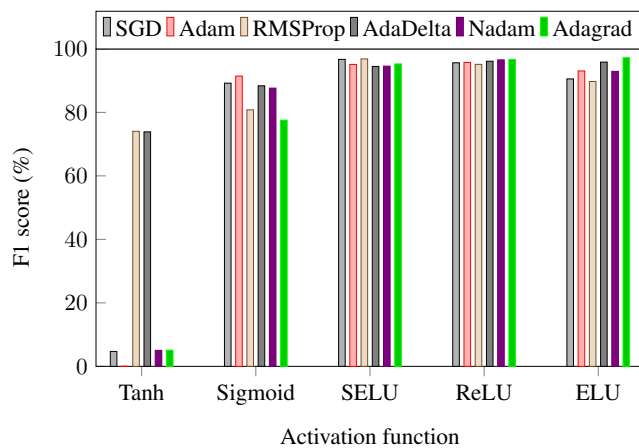**FIGURE 6.** Performance in terms of F1 score for different combinations of activation functions and optimizers.

FIGURE 7. Performance comparison for diverse ratios of the COVID-19 X-ray images generated by the GAN with respect to the existing number of samples in the dataset.

$\alpha = 1.0$ and the optimizer set to Adagrad with the learning rate of 0.001. For this first experimental setting for selecting the optimal hyper-parameters of the deep learning-based model, the mini-batch size (B) was set to 8, and the number of epochs (ξ) was set to 20. With this configuration, the validation accuracy, precision, and F1 score were found to be 97.25%, 97.24%, and 97.21%, respectively. Therefore, for further analysis, we applied this configuration in the customized CNN model of our DL-CRC framework. Furthermore, in the max-pooling layer of our proposed CNN architecture, we conducted manual parameter tuning, and the pool size k_i was assigned as μ, where μ=2% of the initial size of the input x_i.

In the second experimental scenario, as the number of COVID-19 samples in the collected dataset was lower than the pre-defined threshold δ, we applied our proposed DARI algorithm to increase the number of COVID-19 samples so that the model can be trained with a robust training data and eventually predict positive COVID-19 cases with high accuracy. In Fig. 7, we altered the proportions for our customized GAN model in the DARI algorithm with respect to the original sample size of the COVID-19 class. The ratios of GAN-generated samples of the proposed approach were varied from 50% to 200% with respect to the number of COVID-19 examples in the original dataset. The number of iterations for producing the augmented samples using the GAN-based method was set to 200. Among the proportions mentioned earlier, the COVID-19 detection performance of our customized CNN model was found to be the highest (with an accuracy of 93.94%) when the number of newly generated samples was 100% of the size of the original COVID-19 samples. Therefore, we picked this configuration to be used in our conducted experiments in the next scenario.

After producing the augmented samples for the COVID-19 class, we analyzed the effect of combining the adaptive generic data augmentation and GAN-based DARI algorithm with the CNN architecture to fully implement and fine-tune the DL-CRC framework, and compared the performance with

model of our proposed DL-CRC framework and carried out a grid search to achieve the optimal model parameters (i.e., the best activation functions and optimizer). It is worth noting that other customized CNN models revealed a performance bottleneck in terms of validation accuracy and we found the model in Fig. 3 to be the most lightweight yet efficient for automating the chest X-ray classification task. Figs. 4, 5, and 6 demonstrate the results obtained from the hyper-parameter tuning in terms of accuracy, precision, and F1 score, respectively. These performances were extensively evaluated across six optimizers (Stochastic Gradient Descent (SGD), Adaptive Moment Estimation (Adam), Root Mean Square Propagation (RMSProp), Adaptive Delta (AdaDelta), Nesterov and Adam (Nadam), and Adaptive Gradient Algorithm (Adagrad)) and five activation functions (tanh, sigmoid, Scaled Exponential Linear Unit (SELU), Rectified Linear Unit (ReLU), and Exponential Linear Unit (ELU)). As depicted by the results in these figures, SELU demonstrated better performances on average when compared with the other activation functions. However, the best performance was exhibited when ELU is adopted as the activation function with the value of constant
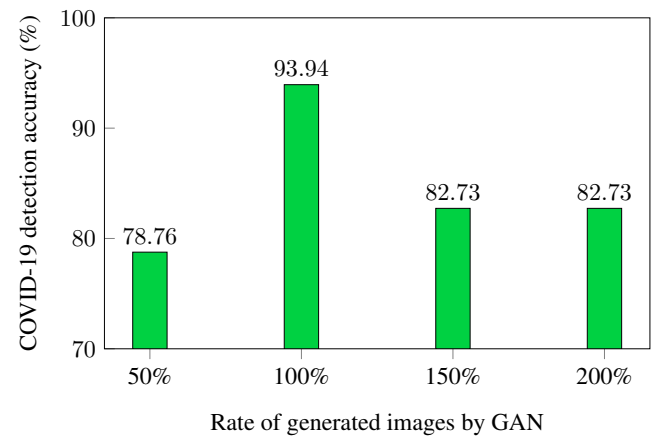
IEEE Access

(a) Proposed CNN without data augmentation.

(b) Proposed CNN and GAN-based DL-CRC (for efficient COVID-19 detection accuracy).
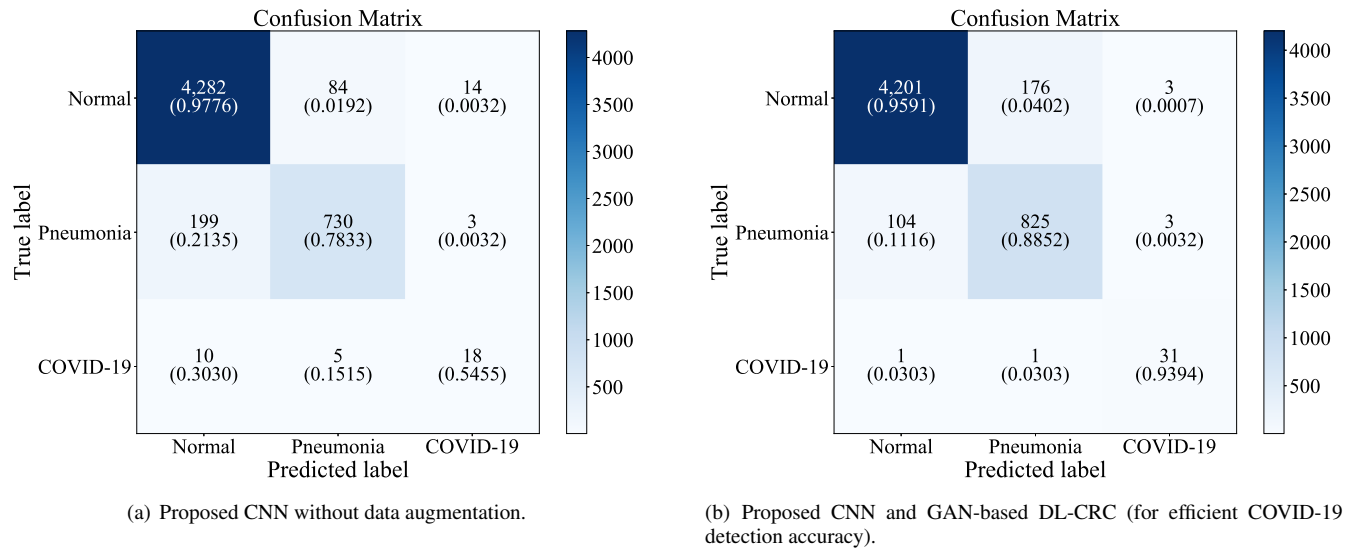
FIGURE 8. Confusion matrix of testing phase employing 5-fold stratified cross-validation.

the base CNN model only (i.e., without adopting DARI algorithm). The experiment was conducted utilizing a five-fold stratified cross-validation. Using the stratification technique, the samples are rearranged so that each fold has a stable representation of the whole dataset by maintaining the percentage of samples for each class [33]. In our third experimental setup, the number of epochs (ξ) was set to 100, and the mini-batch size (B) was set to 8. The number of convolutional layers, n_c, was set to five. The number of fully-connected/dense layers, n_d, was also fixed to five. Note that these hyperparameter values were manually tuned. To analyze the results more critically in terms of COVID-19 detection efficiency, in this experimental setting, we also investigated the normalized and non-normalized values of the confusion matrices of our customized CNN model without (i.e., CNN-only model) and with the proposed DARI algorithm (i.e., the complete DL-CRC framework). Fig. 8(a) represents the normalized confusion matrix where the proposed CNN model is implemented without applying the data augmentation, and Fig. 8(b) depicts the same for the combined CNN and DARI algorithm. Despite similar performances of both approaches, the normalized confusion matrix demonstrates that our proposed DL-CRC framework is much more robust for classifying positive COVID-19 and pneumonia cases. The proposed DL-CRC exhibited 93.94% and 88.52% accuracies while detecting positive COVID-19 and pneumonia cases, respectively. The encouraging classification performance indicates that our proposed deep learning-based DL-CRC framework is able to classify the radiograph images with high efficiency, specifically for COVID-19 detection.

Furthermore, we analyzed the impact of generic and GAN-based data augmentation separately combined with our customized CNN model and compared the COVID-19 detection accuracy with the proposed DL-CRC framework. Table 2 exhibits the simulation results, which proves that both the generic and GAN-based data augmentation had significant

influence in enhancing the COVID-19 detection efficiency. The simulation results in the table show that our CNN-only base model achieved 54.5%, CNN with generic data augmentation obtained 63.4%, and CNN with the proposed GAN-based data augmentation delivered 84.5% COVID-19 detection accuracy. On the other hand, the proposed DL-CRC framework demonstrated the highest COVID-19 detection accuracy (93.94%). This good performance is attributed to the combination of our customized CNN model with the proposed DARI algorithm where both generic and GAN-based data augmentation are adaptively performed, Therefore, it is evident from these results that our proposed DL-CRC framework made the customized CNN model much more robust with DARI algorithm.

In the third experimental scenario, we compared the performance of our customized CNN model with the performances of the state-of-the-art CNN models such as Inception-Resnet V2, Resnet, and DenseNet. The reason behind choosing these contemporary models is their good performances reported in the recent literature for COVID-19 detection. It is worth noting that Inception-ResNet v2 and DenseNet belong to the depth-based and multi-path-based CNN paradigms, respectively. On the other hand, ResNet combines both depth-based and multi-path-based CNN architectures. Table 3 demonstrates the comparative

TABLE 2. Performance comparison of the proposed DL-CRC and CNN with generic and GAN-based data augmentation.

Model	Normal	Pneumonia	COVID-19
CNN-only	97.76%	78.33%	54.55%
CNN & generic	80.48%	23.07%	63.64%
CNN & GAN	88.33%	83.16%	84.84%
Proposed DL-CRC (CNN, GAN, generic)	95.91%	88.52%	93.94%

analysis, which indicates the efficiency of our proposed DL-CRC framework in terms of COVID-19 and pneumonia detection using chest X-ray images. Our proposed model, outperformed ResNet, Inception-ResNet v2, and DenseNet. Although Densenet achieves 98.01% prediction performance for normal test cases, its accuracy is only 72.42% for pneumonia detection while it exhibits the poorest performance of 60.61% for identifying COVID-19 cases. This implies that multi-path-based structure, although reported in recent work, is not suitable for COVID-19 detection. On the other hand, Inception ResNet v2, using the depth-based CNN modeling paradigm, achieves improved COVID-19 detection accuracy (69.70%). The combination of these two modeling paradigms is incorporated in ResNet, which is able to predict test cases having COVID-19 samples slightly elevated accuracy of 72.72%. On the other hand, our proposed DL-CRC framework, combining our envisioned DARI algorithm and customized CNN model, is able to detect the COVID-19 cases with a significantly high accuracy of 93.94%. Note that the pneumonia (the other abnormal case) present in the test dataset is also detected with much higher accuracy (88.52%) compared to the contemporary models. Even though the performance slightly drops for normal case identification, the accuracy is still close to 96% in case of our proposal. Furthermore, in the final column of Table 3, the AUC (area under the ROC (receiver operating characteristic) curve) values are also listed for the proposed DL-CRC and contemporary models. The AUC score of our proposed DL-CRC is 0.9525 which demonstrates the reasonable accuracy of identification across all samples in the test data. Thus, the encouraging performance of the proposed DL-CRC algorithm over prominent CNN models clearly demonstrates that the proposed technique can be useful for detecting COVID-19 and pneumonia cases with a significantly high (i.e., reliable) accuracy.

Furthermore, we compare the performance of our proposal with a recent custom model, referred to as Dark-CovidNet [18]. For multi-class classification, the accuracy of DarkCovidNet was reported to be 87.02%, which is considerably lower than that of our proposed model's performance (93.94%), which we believe ensures the effectiveness of our proposed model. In addition, we have conducted two-fold experiments to validate and compare our proposed technique (DL-CRC) with DarkCovidNet. Table 4 demonstrates the results obtained when our proposed model is tested on both

datasets, and the DarkCovidNet model is tested on both datasets. Both models were trained by employing the respective dataset used by the work in [18] and our current work. These experimental results presented in Table 4 were produced after training the models for 25 epochs for each case, and then the trained models were tested on both datasets. Our proposed technique outperformed DarkCovidNet for detection accuracies for both normal and COVID-19 cases. In addition to the classification efficiency, our proposed DL-CRC framework is more lightweight than that of used in DarkCovidNet. Our customized CNN model of DL-CRC consists of 5 convolutional layers while the DarkCovid-Net model comprises 17 convolutional layers, making our model's training phase more lightweight and computationally less expensive than the DarkCovidNet model.

Moreover, while some researches reported overall accuracy, they did not mention the COVID-19 detection accuracy. On the other hand, most researches applying deep learning techniques did not report the AUC score, which is a robust representative performance metric for practically evaluating the COVID-19 detection ability of the model. In summary, by applying various contemporary CNN models (Inception with Resenet V2, Resnet, Densenet) and a recent customized model (DarkCovidNet) for COVID-19 detection on the latest dataset compiled from four public repositories, we realized that their reported performances are constrained by overfitting and influenced by biased test data. Thus, the accuracy bottleneck of those existing models justifies why we required to build a customized CNN model in this research and combine it with the DARI algorithm to perform robust training and avoid overfitting to ensure high COVID-19 detection accuracy and a significantly high AUC score.

VI. LIMITATIONS OF THE STUDY

In this section, we briefly discuss some limitations and possible future work that can be conducted to extend the study.

- Our study and experiments have been conducted at a very critical stage and time-sensitive manner to combat the COVID-19 pandemic with a proof-of-concept COVID-19 using radiograph images. Despite compiling datasets from multiple sources with X-ray images containing COVID-19 samples, the used data was considerably small in size. Therefore, synthetic images were generated using our customized GAN-assisted data augmentation technique that were used to train a robust CNN model to perform binary (normal and COVID-19) and three-way classification (normal, pneumonia, and COVID-19) with significantly high accuracy. Due to the lack of real datasets consisting of other diseases (e.g., SARS, MERS, and so forth) which exhibit acute respiratory distress syndrome (ARDS) and pneumonia-like conditions in the lungs, more class labels were not considered in our work.
- From a physician's perspective, it is important to diagnose the severity of COVID-19. However, due to the lack of labeled data, in this work, our model could not

TABLE 3. Performance comparison of our proposed DL-CRC architecture with the existing CNN architectures for all three classes.

Model	Normal	Pneumonia	COVID-19	AUC Score
Proposed DL-CRC	95.91%	88.52%	93.94%	0.9525
Inception-ResNet V2	97.06%	77.04%	69.70%	0.9473
ResNet	98.83%	74.67%	72.72%	0.9314
DenseNet	98.01%	72.42%	60.61%	0.8798

TABLE 4. Comparison of the performance our proposed model with that of DarkCovidNet [18] on both datasets.

Model	Test dataset	Normal	Pneumonia	COVID-19
DL-CRC	Dataset used in our paper	99.45%	84.84%	83.04%
	Dataset in DarkCovidNet paper	93%.00	6.00%	80.00%
DarkCovidNet	Dataset used in our paper	13.33%	82.94%	45.00%
	Dataset in DarkCovidNet paper	84.00%	80.00%	76.00%

be used to classify the various stages of COVID-19 such as asymptomatic, mild, high and severe.

- The proposed technique performed efficiently when we utilized it to analyze X-ray samples. However, the study can be extended to evaluate the system's performance in COVID-19 detection while using other radiograph techniques such as CT scan, lung ultrasound, and lung PET (positron emission tomography) scan.

- The dataset used in this study is limited by only one modality type, i.e., X-ray images containing COVID-19 features. Further customization in our CNN model will be required if we want to combine multiple imaging modalities (e.g., lung CT scan, ultrasound, PET along with X-ray images), other modalities (e.g., body temperature, ECG, MCG, diabetes level, renal function, and so forth), and patient parameters (e.g., age, gender, ethnicity, travel history, and contact history) to perform an in-depth COVID-19 classification. Therefore, a multi-modal input characterization and corresponding AI model customization will be needed in the future for interpreting and explaining the classification results.

VII. CONCLUSION

In this paper, we addressed the emerging challenges of detecting COVID-19. Due to the shortage of efficient diagnosis equipment and personnel in many areas, particularly in developing and/or rural zones, numerous people remain non-diagnosed. This results in a substantial gap between the number of confirmed and actual cases. Radiographs such as chest X-ray images and CT scans have been demonstrated to have the potential for detecting COVID-19 infection in the lungs that can complement the time-consuming viral and antibody testing. While CT scans have higher resolution or fine-grained details compared to X-ray images, X-ray machines are pervasive in hospital emergency rooms, public health facilities, and even rural health centers or clinics. In addition, because X-ray is a much cheaper alternative and an appealing solution for portability in mobile trucks and COVID-19 screening booths with adequate shielding and power supply, how to identify COVID-19 infection of the lung by recognizing patterns such as glass opacities and lung consolidations raised a formidable research problem, that we addressed in this paper. Also, we discussed why it is necessary to automate the X-ray image classification to be well prepared for the next wave of COVID-19 pandemic, when radiologists and caregivers are expected to be overwhelmed by patient influx as well as the need to self-

isolate in case they themselves become infected. This means there is a pressing need to automate the classification of radiographs, particularly X-ray images, to minimize the turn-around time for COVID-19 detection. Therefore, to leverage the availability and cost-efficiency of chest X-ray imaging, in this paper, we proposed a framework called DL-CRC (Deep learning-based chest radiograph classification) to automate COVID-19 detection that can complement existing viral and antibody testing methods.

Our proposed DL-CRC framework consists of two parts: the DARI algorithm (which adaptively employs a customized generative adversarial network and generic data augmentation techniques such as zoom and rotation) and a two-dimensional convolutional neural network (CNN) model. We employed a unique dataset for multiple publicly available sources, containing radiograph images of COVID-19 and pneumonia infected lungs, along with normal lung imaging. The classification accuracy significantly increased to 94.61% by adopting our proposed DL-CRC framework. Our proposal was compared with existing deep learning models from diverse categories such as depth-based CNN (e.g., Inception-ResNet v2), multi-path-based CNN (DenseNet), and hybrid CNN (ResNet) architectures. Extensive experimental results demonstrated that our proposed combination of DARI and custom CNN-based DL-CRC framework significantly outperformed the existing architectures. Thus, incorporating our proposed model with significantly high accuracy into the clinical-grade as well as portable X-ray equipment can allow an automated and accurate detection of COVID-19 in the scrutinized patients.

REFERENCES

[1] S. Karanam, R. Li, F. Yang, W. Hu, T. Chen, and Z. Wu, "Towards contactless patient positioning," *IEEE Transactions on Medical Imaging*, vol. 39, no. 8, pp. 2701–2710, Aug. 2020.

[2] "WHO coronavirus disease (COVID-19) dashboard." World Health Organization (WHO). https://covid19.who.int/ (accessed Jul. 15, 2020).

[3] W. Wang *et al.*, "Detection of SARS-CoV-2 in different types of clinical specimens," *Journal of the American Medical Association (JAMA)*, vol. 323, no. 18, pp. 1843–1844, May 2020, doi:10.1001/jama.2020.3786.

[4] G. D. Rubin *et al.*, "The role of chest imaging in patient management during the COVID-19 pandemic: A multinational consensus statement from the Fleischner society," *Radiology*, vol. 296, no. 1, pp. 172–180, Jul. 2020, pMID: 32255413, doi: 10.1148/radiol.2020201365.

[5] M. Hosseiny, S. Kooraki, A. Gholamrezanezhad, S. Reddy, and L. Myers, "Radiology perspective of coronavirus disease 2019 (COVID-19): Lessons from severe acute respiratory syndrome and middle east respiratory syndrome," *American Journal of Roentgenology*, vol. 214, no. 5, pp. 1078–1082, May 2020, doi: 10.2214/AJR.20.22969.

[6] "ACR recommendations for the use of chest radiography and computed tomography (CT) for suspected COVID-19 infection." American College of Radiology (ACR). https://www.acr.org/Advocacy-and-Economics/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI
10.1109/ACCESS.2020.3025010, IEEE Access

IEEE Access·

Sakib *et al.*: DL-CRC: Deep Learning-based Chest Radiograph Classification for COVID-19 Detection: A Novel Approach

ACR-Position-Statements/Recommendations-for-Chest-Radiography-and-CT-for-Suspected-COVID19-Infection (accessed Jul. 15, 2020).

[7] T. Ai *et al.*, "Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: A report of 1014 cases," *Radiology*, vol. 296, no. 2, pp. E32–E40, Aug. 2020, pMID: 32101510, doi: 10.1148/radiol.2020200642.

[8] A. Dangis *et al.*, "Accuracy and reproducibility of low-dose submillisievert chest CT for the diagnosis of COVID-19," *Radiology: Cardiothoracic Imaging*, to be published, doi: 10.1148/ryct.2020200196.

[9] A. Jacobi, M. Chung, A. Bernheim, and C. Eber, "Portable chest X-ray in coronavirus disease-19 (COVID-19): A pictorial review," *Clinical Imaging*, vol. 64, pp. 35–42, Aug. 2020, doi: 10.1016/j.clinimag.2020.04.001.

[10] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artificial Intelligence Review*, to be published, doi: 10.1007/s10462-020-09825-6.

[11] M. Mossa-Basha, C. C. Meltzer, D. C. Kim, M. J. Tuite, K. P. Kolli, and B. S. Tan, "Radiology department preparedness for COVID-19: radiology scientific expert review panel," *Radiology*, vol. 296, no. 2, pp. E106–E112, Aug. 2020, pMID: 32175814, doi: 10.1148/radiol.2020200988.

[12] "How does COVID-19 appear in the lungs?" Imaging Technology News. https://www.itnonline.com/content/how-does-covid-19-appear-lungs (accessed Jul. 15, 2020).

[13] R. Vaishya, M. Javaid, I. H. Khan, and A. Haleem, "Artificial Intelligence (AI) applications for COVID-19 pandemic," *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, vol. 14, no. 4, pp. 337–339, Jul.-Aug. 2020, doi: 10.1016/j.dsx.2020.04.012.

[14] J. Zhang, Y. Xie, Y. Li, C. Shen, and Y. Xia, "COVID-19 screening on chest X-ray images using deep learning based anomaly detection," *arXiv preprint arXiv:2003.12338*, Mar. 2020.

[15] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks," *arXiv preprint arXiv:2003.10849*, Mar. 2020.

[16] A. Abbas, M. M. Abdelsamea, and M. M. Gaber, "Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network," *arXiv preprint arXiv:2003.13815*, May 2020.

[17] L. Wang and A. Wong, "COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images," *arXiv preprint arXiv:2003.09871*, May 2020.

[18] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. Rajendra Acharya, "Automated detection of COVID-19 cases using deep neural networks with X-ray images," *Computers in Biology and Medicine*, vol. 121, Jun. 2020, Art. no. 103792, doi: 10.1016/j.compbiomed.2020.103792.

[19] X. Xu *et al.*, "Deep learning system to screen coronavirus disease 2019 pneumonia," *arXiv preprint arXiv:2002.09334*, Feb. 2020.

[20] O. Gozes, M. Frid-Adar, H. Greenspan, P. D. Browning, H. Zhang, W. Ji, A. Bernheim, and E. Siegel, "Rapid AI development cycle for the coronavirus (COVID-19) pandemic: initial results for automated detection & patient monitoring using deep learning CT image analysis," *arXiv preprint arXiv:2003.05037*, Mar. 2020.

[21] A. A. Ardakani, A. R. Kanafi, U. R. Acharya, N. Khadem, and A. Mohammadi, "Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks," *Computers in Biology and Medicine*, vol. 121, Jun. 2020, Art. no. 103795, doi: 10.1016/j.compbiomed.2020.103795.

[22] T. Iqbal and H. Ali, "Generative adversarial network for medical images (MI-GAN)," *Journal of Medical Systems*, vol. 42, no. 11, Oct. 2018, Art. no. 231, doi: 10.1007/s10916-018-1072-9.

[23] X. Yi, E. Walia, and P. Babyn, "Generative adversarial network in medical imaging: A review," *Medical Image Analysis*, vol. 58, Dec. 2019, Art. no. 101552, doi: 10.1016/j.media.2019.101552.

[24] M. Chung *et al.*, "CT imaging features of 2019 novel coronavirus (2019-nCoV)," *Radiology*, vol. 295, no. 1, pp. 202–207, Apr. 2020, pMID: 32017661, doi: 10.1148/radiol.2020200230.

[25] I. Goodfellow *et al.*, "Generative adversarial nets," *arXiv preprint arXiv:1406.2661*, Jun. 2014.

[26] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," presented at the *Workshop on Deep Learning for Audio, Speech and Language Processing, International Conference on Machine Learning (ICML)*, vol. 30, Atlanta, GA, USA, Jun. 16–21, 2013.

[27] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 10, pp. 6232–6251, Oct. 2016, doi: 10.1109/TGRS.2016.2584107.

[28] W. S. Kim, H. Jung, and J. H. Choi, "Automatic CAC Voxel classification with multi-scale CNN architecture," in *Proc. International Conference on Information and Communication Technology Convergence (ICTC)*, Jeju Island, Korea (South), Oct. 16–18, 2019, pp. 1351–1353, doi: 10.1109/ICTC46691.2019.8939821.

[29] J. P. Cohen, P. Morrison, and L. Dao, "COVID-19 image data collection," *arXiv 2003.11597*, Mar. 2020.

[30] D. S. Kermany *et al.*, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, Feb. 2018, doi: 10.1016/j.cell.2018.02.010.

[31] J. Irvin *et al.*, "CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison," in *Proc. AAAI Conference on Artificial Intelligence*, vol. 33, Jul. 2019, pp. 590–597, doi: 10.1609/aaai.v33i01.3301590.

[32] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. Summers, "ChestX-Ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, Jul. 21–26, 2017, pp. 3462–3471, doi: 10.1109/CVPR.2017.369.

[33] R. Kohavi, "A Study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. 14th International Joint Conferences on Artificial Intelligence (IJCAI)*, vol. 2, Aug. 1995, pp. 1137–1143.

· · ·