

Convolutional Sparse Support Estimator-Based COVID-19 Recognition From X-Ray Images

Mehmet Yamaç¹, Mete Ahishali¹, Aysen Degerli¹, Serkan Kiranyaz², *Senior Member, IEEE*,
Muhammad E. H. Chowdhury³, *Senior Member, IEEE*, and Moncef Gabbouj⁴, *Fellow, IEEE*

Abstract—Coronavirus disease (COVID-19) has been the main agenda of the whole world ever since it came into sight. X-ray imaging is a common and easily accessible tool that has great potential for COVID-19 diagnosis and prognosis. Deep learning techniques can generally provide state-of-the-art performance in many classification tasks when trained properly over large data sets. However, data scarcity can be a crucial obstacle when using them for COVID-19 detection. Alternative approaches such as representation-based classification [collaborative or sparse representation (SR)] might provide satisfactory performance with limited size data sets, but they generally fall short in performance or speed compared to the neural network (NN)-based methods. To address this deficiency, convolution support estimation network (CSEN) has recently been proposed as a bridge between representation-based and NN approaches by providing a noniterative real-time mapping from query sample to ideally SR coefficient support, which is critical information for class decision in representation-based techniques. The main premises of this study can be summarized as follows: 1) A benchmark X-ray data set, namely QaTa-Cov19, containing over 6200 X-ray images is created. The data set covering 462 X-ray images from COVID-19 patients along with three other classes; bacterial pneumonia, viral pneumonia, and normal. 2) The proposed CSEN-based classification scheme equipped with feature extraction from state-of-the-art deep NN solution for X-ray images, CheXNet, achieves over 98% sensitivity and over 95% specificity for COVID-19 recognition directly from raw X-ray images when the average performance of 5-fold cross validation over QaTa-Cov19 data set is calculated. 3) Having such an elegant COVID-19 assistive diagnosis performance, this study further provides evidence that COVID-19 induces a unique pattern in X-rays that can be discriminated with high accuracy.

Index Terms—Coronavirus disease (COVID-19) recognition, representation-based classification, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus, transfer learning.

I. INTRODUCTION

CORONAVIRUS disease 2019 (COVID-19) has been declared as a pandemic by the World Health Organization (WHO) a few months after its first appearance. It has infected more than 70 million people, caused a few million casualties, and has so far paralyzed mobility all around the

world. The spreading rate of COVID-19 is so high that the number of cases is expected to be doubled every three days if the social distancing is not strictly observed to slow this accretion [1]. Roughly around half of the COVID-19 positive patients also exhibit a comorbidity [2], making it difficult to differentiate COVID-19 from other lung diseases. Automated and accurate COVID-19 diagnosis is critical for both saving lives and preventing its rapid spread in the community. Currently, reverse transcription-polymerase chain reaction (RT-PCR) and computed tomography (CT) are the common diagnostic techniques used today. RT-PCR results are ready at the earliest 24 h for critical cases and generally take several days to conclude a decision [3]. CT may be an alternative at initial presentation; however, it is expensive and not easily accessible [4]. The most common tool that medical experts use for both diagnostic and monitoring the course of the disease is X-ray imaging. Compared to RT-PCR or CT test, having an X-ray image is an extremely low cost and a fast process, usually taking only a few seconds. Recently, WHO reported that even RT-PCR may give false results in COVID-19 cases due to several reasons such as poor quality specimen from the patient, inappropriate processing of the specimen, taking the specimen at an early or late stage of the disease [5]. For this reason, X-ray imaging has a great potential to be an alternative technological tool to be used along with the other tests for an accurate diagnosis.

In this study, we aim to differentiate X-ray images of COVID-19 patients among other classes; bacterial pneumonia, viral pneumonia, and normal. For this work, a benchmark COVID-19 X-ray data set, Qata-Cov19 (Qatar University and Tampere University COVID-19 Data set) that contains 462 X-ray images from COVID-19 patients was collected. The images in the data set are different in quality, resolution, and SNR levels as shown in Fig. 1. QaTa-Cov19 also contains many X-ray images from the COVID-19 patients who are in the early stages; therefore, their X-ray images show mild or no-sign of COVID-19 infestation by the naked eye.¹ Some sample images are shown in Fig. 2(b). Another fact that makes the diagnosis far more challenging is that interclass similarity can be very high for many X-ray images as some samples are shown in Fig. 2(a). Against such high interclass similarities and intraclass variations, in this study, we aim for a high robustness level.

In numerous classification tasks, deep learning techniques have been shown to achieve state-of-the-art performance in

Manuscript received May 7, 2020; revised October 9, 2020 and December 21, 2020; accepted March 23, 2021. Date of publication April 19, 2021; date of current version May 3, 2021. (Corresponding author: Mehmet Yamaç.)

Mehmet Yamaç, Mete Ahishali, Aysen Degerli, and Moncef Gabbouj are with the Faculty of Information Technology and Communication Sciences, Tampere University, 33720 Tampere, Finland (e-mail: mehmet.yamac@tuni.fi).

Serkan Kiranyaz and Muhammad E. H. Chowdhury are with the Department of Electrical Engineering, Qatar University, Doha 2713, Qatar.

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TNNLS.2021.3070467>.

Digital Object Identifier 10.1109/TNNLS.2021.3070467

¹The statements belong to the medical doctors whose names are listed in the Acknowledgment section.

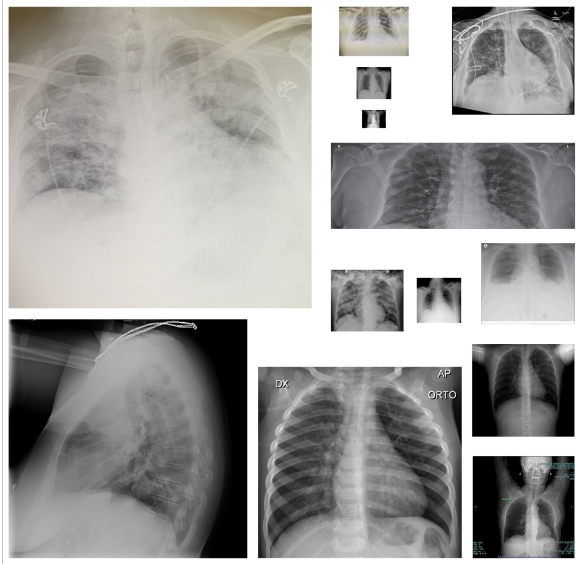


Fig. 1. Sample COVID-19 X-ray images from QaTa-Cov19.

terms of both recognition accuracy and their parallelizable computing structures which play an important role, especially in real-time applications. Despite their advantages, in order to achieve the desired performance level in a deep model, proper training over a massive training data set is usually needed. Nevertheless, this is unfortunately unfeasible for this problem since the available data is still rather limited.

An alternative supervised approach, which requires a limited number of training samples to achieve satisfactory classification accuracy is representation-based classification [6]–[8]. In representation-based classification systems, a dictionary, the columns of which consist of the training samples that are stacked in such a way that a subset of them corresponding to a class, is predefined. A test sample is expected to be a linear combination of all points from the same class as the test sample. Therefore, given a predefined dictionary matrix, \mathbf{D} and a test sample \mathbf{y} , we expect the solution $\hat{\mathbf{x}}$ from $\mathbf{y} = \mathbf{D}\mathbf{x}$, carry enough information about the class of \mathbf{y} . Overall, in this study, we draw a convolutional support estimation network (CSEN) [9]-based solution pipeline, which fuses the representation-based classification scheme into a neural network (NN) body.

The rest of this article is organized as follows. In Section II, notations and mathematical preliminaries are given with emphasis on sparse representation (SR) and sparse support estimation (SE). Then in Section III, a literature review on deep learning models over X-ray images and representation-based classification is presented. The proposed CSEN-based COVID-19 recognition system is introduced in Section IV along with two recent alternative approaches that are used as the competing methods. The data collection is also explained in this section. Experimental setup and the main results are provided in Section V. Finally, Section VII concludes this article and suggests topics for future research.

II. PRELIMINARIES AND MATHEMATICAL NOTATIONS

A. Notations

In this study, the ℓ_p -norm of a vector $\mathbf{x} \in \mathbb{R}^n$ is defined as $\|\mathbf{x}\|_{\ell_p} = (\sum_{i=1}^n |x_i|^p)^{1/p}$ for $p \geq 1$. On the other hand,

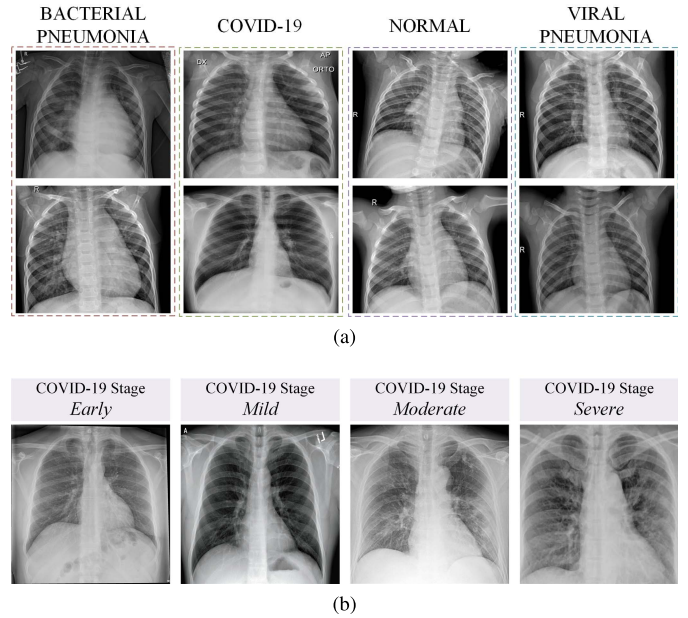


Fig. 2. Sample QaTa-Cov19 X-ray images. (a) X-ray images from different classes. (b) X-ray images from the COVID-19 patients who are in the different stages.

the ℓ_0 -norm of the vector $\mathbf{x} \in \mathbb{R}^n$ is defined as $\|\mathbf{x}\|_{\ell_0} = \lim_{p \rightarrow 0} \sum_{i=1}^n |x_i|^p = \#\{j : x_j \neq 0\}$ and the ℓ_∞ -norm is defined as $\|\mathbf{x}\|_{\ell_\infty} = \max_{i=1, \dots, n} (|x_i|)$. A signal \mathbf{s} is called strictly k -sparse if $\|\mathbf{x}\|_0 \leq k$. Sparse support set or simply support set, $\Lambda \subset \{1, 2, 3, \dots, n\}$ of sparse signal \mathbf{x} can be defined as the set of nonzero coefficients' location, i.e., $\Lambda := \{i : x_i \neq 0\}$.

B. Sparse Signal Representation

SR of a signal $\mathbf{s} \in \mathbb{R}^d$ in a predefined set of waveforms, $\Phi \in \mathbb{R}^{d \times n}$, can be defined as representing \mathbf{s} as a linear combination of only a small subset of atoms in the dictionary Φ , i.e., $\mathbf{s} = \Phi\mathbf{x}$. Defining these sets, which dates back to Fourier's pioneering work [10], has been excessively studied in the literature. In the early approaches, these sets of waveforms have been selected as a collection of linearly independent and generally orthogonal waveforms (which are called a complete dictionary or basis, i.e., $d = n$) such as Fourier transform, DCT, and wavelet transform, until the pioneering work of Mallat [11] on overcomplete dictionaries ($n \gg d$). In the last decade, interest in SR research increased tremendously. Their wide range of applications includes denoising [12], classification [13], anomaly detection [14], [15], deep learning [16], and compressive sensing (CS) [17], [18].

With a possible dimensional reduction that can be satisfied via a compression matrix $\mathbf{A} \in \mathbb{R}^{m \times d}$ ($m \ll d$), sample can be obtained from \mathbf{s}

$$\mathbf{y} = \mathbf{A}\mathbf{s} = \mathbf{A}\Phi\mathbf{x} = \mathbf{D}\mathbf{x} \quad (1)$$

where $\mathbf{D} \in \mathbb{R}^{m \times n}$ can be called the equivalent dictionary. Because (1) describes an underdetermined system of linear equations, finding the representation coefficient vector \mathbf{x} requires at least one more constraint to have a unique solution. Using the prior information about sparsity, the following

representation:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \text{ s.t. } \mathbf{D}\mathbf{x} = \mathbf{y} \quad (2)$$

which is also an SR of \mathbf{x} has a unique solution provided that \mathbf{x} is strictly sparse and \mathbf{D} satisfies some required properties [19]. For instance, if $\|\mathbf{x}\|_0 = k$, the minimum number of linearly independent columns of \mathbf{D} , $\text{spark}(\mathbf{D})$, should be greater than $2k$, i.e., $\text{spark}(\mathbf{D}) \geq 2k$ in order to not to have $\mathbf{D}\mathbf{x}' = \mathbf{D}\mathbf{x}''$ for distinct k -sparse signals, \mathbf{x}' and \mathbf{x}'' [19]. However, the optimization problem in (2) is a NP-hard. Fortunately, the following relaxation:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \text{ s.t. } \mathbf{D}\mathbf{x} = \mathbf{y} \quad (3)$$

produces exactly the same solution as that of (2) provided that \mathbf{D} obeys some criteria: the equivalence of ℓ_0 - ℓ_1 minimization problems can be guaranteed when \mathbf{D} satisfies a notation of null space property (NSP) [20], [21] not only for exact sparse signals but approximately sparse signals. Furthermore, the query sample \mathbf{y} can be corrupted with an additive noise pattern. In this case, the equality constraint in (3) can be further relaxed such as in the basis pursuit denoising (BPDN) [22]: $\min_{\mathbf{x}} \|\mathbf{x}\|$ s.t. $\|\mathbf{y} - \mathbf{D}\mathbf{x}\| \leq \epsilon$, where ϵ is a small constant that depends on the noise level. In this case, a stronger property which is known as restricted isometry property (RIP) [23], [24] is frequently used which both cover conditions satisfying exact recovery of BP and stable recovery of BPDN, e.g., exact recovery of \mathbf{x} from (3) is possible when \mathbf{D} has RIP and $m > k(\log(n/k))$.

We may refer to the sparse SE problem as finding the indices a set, Λ , of nonzero elements of \mathbf{x} [25], [26]. Indeed, in many applications, SE can be more important than finding the magnitude and sign of \mathbf{x} as well as Λ , which refers to the sparse signal recovery (SSR) via a recovery technique, such as (3). For example, in a sparse representation-based classification (SRC) system, a query sample \mathbf{y} can be represented with sparse coefficient vector, \mathbf{x} , in the dictionary, \mathbf{D} in such a way that when we recover this representation coefficient from $\mathbf{y} = \mathbf{D}\mathbf{x}$, the solution vector $\hat{\mathbf{x}}$ is expected to have a significant number of nonzero coefficients coming from the particular locations corresponding to the class of \mathbf{y} .

Readers are referred to [9] for a more detailed literature review on SE and its applications. In the sequel, we briefly summarize the building blocks of the proposed approach.

III. BACKGROUND AND PRIOR ART

A. CheXNet

In the proposed approach, we first use the pretrained deep network, CheXNet, to extract discriminative features from raw X-ray images. CheXNet was developed for pneumonia detection from the chest X-ray images [27]. In [27], it was claimed that CheXNet can perform even better than expert radiologists in the pneumonia detection problem. This deep NN design is based on the previously proposed DenseNet [28] that consists of 121 layers. It is first pretrained over ImageNet data set [29] and performed transfer learning over 112120 frontal-view chest X-ray images in the ChestX-ray14 data set [30].

B. Representation-Based Classification

Consider we are given a test sample \mathbf{y} , which represents either the extracted features, \mathbf{s} , or their dimensionally reduced version, i.e., $\mathbf{y} = \mathbf{A}\mathbf{s}$. In developing the dictionary, training samples are stacked in the dictionary \mathbf{D} with particular locations in such a way that the optimal support for a given query \mathbf{y} should be the set of all points coming from the same class as \mathbf{y} . Therefore, a solution vector, $\hat{\mathbf{x}}$ of $\mathbf{y} = \mathbf{D}\mathbf{x}$ is supposed to have enough information, i.e., the sparse support should be the set of location indices of the training sample from the same class as \mathbf{y} . This strategy is generally known as representation-based classification. However, a typical solution $\hat{\mathbf{x}}$ of $\mathbf{y} = \mathbf{D}\mathbf{x}$ is not necessarily a sparse one especially when its size grows with more training samples, which results in a highly underdetermined system of linear equations. Fortunately, if one estimates the representation coefficient vector with a sparse recovery design such as ℓ_1 -minimization as in (3), we can expect that the important nonzero entries of the solution, $\hat{\mathbf{x}}$, are grouped in the particular locations that correspond to the locations of the training samples from the same class as \mathbf{y} . This can be a typical example of scenarios where SE can be more valuable than the magnitudes and sign recovery as explained in Section II-B.

For instance, Wright *et al.* [8] proposed a systematic way of determining the identity of face images using ℓ_1 -minimization. The authors develop a three-step classification technique that includes: (i) normalization of all the atoms in \mathbf{D} and \mathbf{y} to have unit ℓ_2 -norm; (ii) estimating the representation coefficient vector via sparse recovery, i.e., $\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_1$ s.t. $\|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2$; and (iii) finding the residuals corresponding to each class via $\mathbf{e}_i = \|\mathbf{y} - \mathbf{D}_i \hat{\mathbf{x}}_i\|_2$, where $\hat{\mathbf{x}}_i$ is the group of the estimated coefficients, $\hat{\mathbf{x}}$, that correspond to class i .

This technique, which is known as SRC, and its variants have been applied to a wide range of applications in the literature [31], [32], e.g., human action recognition [33], and hyperspectral image classification [34], to name a few. Despite the good recognition accuracy performance of SRC systems, their main drawbacks is the fact that their sparse recovery algorithms (e.g., ℓ_1 -minimization) are iterative methods and computationally costly, rendering them infeasible in real-time applications. Later, the authors of [6] introduced collaborative representation-based classification (CRC), which is similar to SRC except for the use of traditional ℓ_2 -minimization in the second step; $\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \{\|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_2^2\}$. Thus, CRC does not require an iterative solution to obtain representation coefficient thanks to that ℓ_2 -minimization has a closed form solution, $\hat{\mathbf{x}} = (\mathbf{D}^T \mathbf{D} + \lambda \mathbf{I}_{n \times n})^{-1} \mathbf{D}^T \mathbf{y}$. Although, the sparsity in $\hat{\mathbf{x}}$ cannot be guaranteed, it has often been reported to achieve a comparable classification performance, especially in small-size training data sets.

IV. PROPOSED APPROACH

For a computer-aided COVID-19 recognition system design, our primary objective is to achieve the highest sensitivity possible in the diagnosis of COVID-19 induced pneumonia with an acceptable false-alarm rate (e.g., specificity > 95%). In particular, the misdiagnosis of a COVID-19 X-ray image

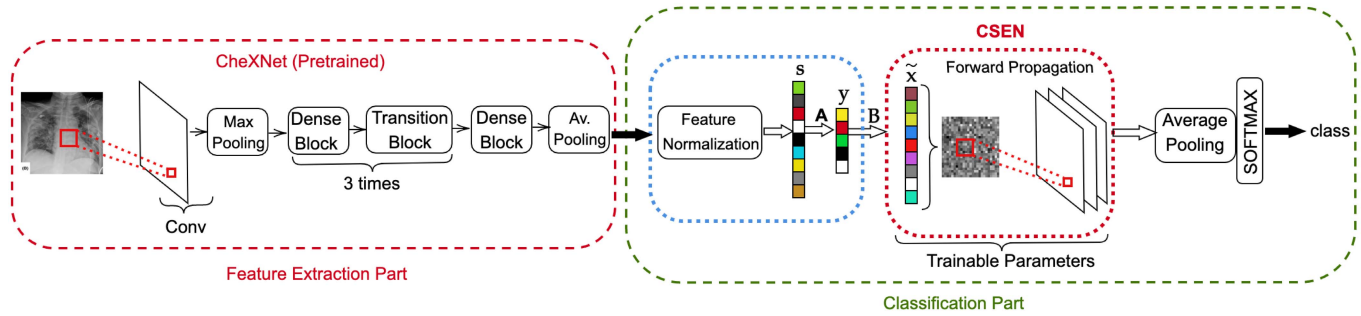


Fig. 3. Proposed approach for Covid recognition from X-ray images. The proposed convolution support estimator network (CSEN) which can be trained from a moderate size training set. The pipeline employs the pretrained deep NN for feature extraction. \mathbf{A} is the dimensional reduction (PCA) matrix, the coarse estimation of representation coefficient (sparse in ideal case), $\hat{\mathbf{x}}$ is obtained via the denoiser matrix, $\mathbf{B} = (\mathbf{D}^T \mathbf{D} + \lambda \mathbf{I})^{-1} \mathbf{D}^T$, where $\mathbf{D} = \mathbf{A} \Phi$ and Φ is the predefined dictionary matrix of training samples (before dimensional reduction).

as a normal case should be minimized whilst a small number of false negatives (FNs) is tolerable.

Our interest in representation-based classification is that they perform well in classification tasks even in the cases where training data is scarce. As mentioned, the two well-known representation-based classification methodologies are SRC [7] and CRC [6]. Among them, SRC provides slightly improved accuracy by solving an SR problem, i.e., producing a sparse solution $\hat{\mathbf{x}}$ from $\mathbf{y} = \mathbf{D}\mathbf{x}$. Then, the location of the nonzero elements of $\hat{\mathbf{x}}$, which is also known as support set, provides the class information of the query \mathbf{y} . Despite improved recognition accuracy, SRC solutions are iterative solutions and can be computationally demanding compared to CRC. In a recent work [9], a compact NN design that can be considered as a bridge between NN-based and representation-based methodologies was proposed. The so-called CSEN uses a predefined dictionary and learns a direct mapping using moderate/low size training set, which maps query samples, \mathbf{y} , directly to the support set of representation coefficients, \mathbf{x} (as it should be purely sparse in the ideal case).

In this study, to address the data scarcity limitations in COVID-19 diagnosis from X-ray images we propose a CSEN-based approach. Since a relatively larger set of COVID-19 X-ray images ever compiled is used in this study, the proposed approach can be evaluated rigorously against a high level of diversity to obtain a reliable analysis. The general pipeline of the proposed CSEN-based recognition scheme is illustrated in Fig. 3. In order to obtain highly discriminative features, we use the recently proposed CheXNet [27], which is the fine-tuned version of 121 layer Dense Convolutional Network (DenseNet-121) [28] by using over 100000 frontal view X-ray images from 14 classes. Having the pretrained CheXNet for feature extraction, we develop two different strategies to obtain the classes of query X-ray images: 1) using CRC with proper preprocessing; 2) a slightly modified version of our recently proposed convolution support estimator (CSEN) models. In the sequel, both techniques will be explained in detail as well as alternative solutions.

A. Benchmark Data Set: QaTa-Cov19

Accordingly, there are several recent works [35]–[38] that have been proposed for COVID-19 detection/classification from X-ray images. However, they use a rather small data set

(the largest containing only a few hundreds of X-ray images), with only a few COVID-19 samples. This makes it difficult to generalize their results in practice. To address this deficiency and provide reliable results, in this study the researchers of Qatar University and Tampere University have compiled a benchmark Covid-19 data set, called QaTa-Cov19. Compared to the earlier benchmark data set created in this domain, such as COVID Chestxray Data set [39] or COVID-19 DATA SET [40], QaTa-Cov19 has the following unique benchmarking properties. First, it is a larger data set, not only in terms of the number of images (more than 6200 images) but its versatility, i.e., QaTa-Cov19 contains additional major pneumonia categories, such as viral and bacterial, along with the control (normal) class. Moreover, this is a diverse data set encapsulating X-ray images from several countries (e.g., Italy, Spain, China, etc.) produced by different X-ray machines.

COVID-19 chest X-ray images were gathered from different publicly available but scattered image sources. However, the major sources of COVID-19 images are Italian Society of Medical and Interventional Radiology (SIRM) COVID-19 Database [40], Radiopaedia [41], Chest Imaging (Spain) at thread reader [42] and online articles and news portals [43]. The authors have carried out the task of collecting and indexing the X-ray images for COVID-19 positive cases reported in the published and preprint articles from China, South Korea, USA, Taiwan, Spain, and Italy, as well as online news-portals (up to 20th April 2020). Therefore, these X-ray images represent different age groups, gender, ethnicity, and country. Negative Covid19 cases were normal, viral, and bacterial pneumonia chest X-ray images and collected from the Kaggle chest X-ray database. Kaggle chest X-ray database contains 5863 chest X-ray images of normal, viral, and bacterial pneumonia with varying resolutions [44]. Out of these 5863 chest X-ray images, 1583 images are normal images and the remaining are bacterial and viral pneumonia images. Sample X-ray images from QaTa-Cov19 data set are shown in Fig. 4.

B. Feature Extraction

With their outstanding performance in image classification along with other inference tasks, deep NNs became a dominant paradigm. However, these techniques usually necessitate a large number of training samples (e.g., several

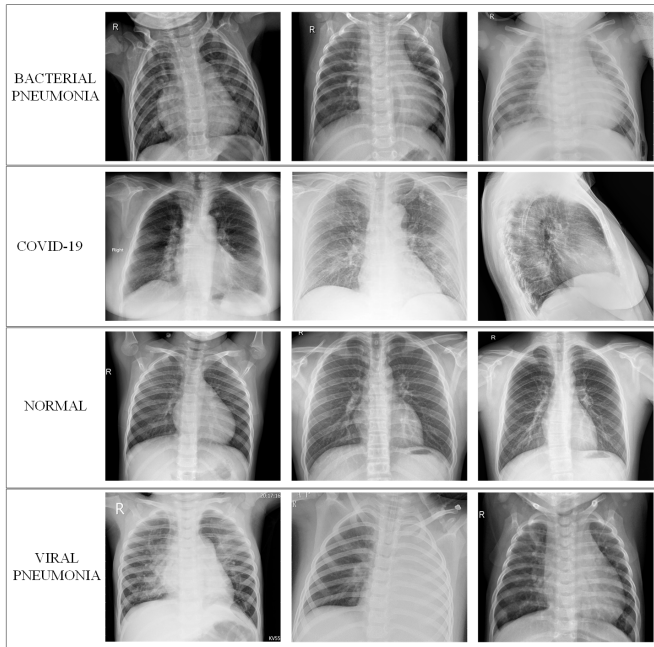


Fig. 4. Samples from the benchmark QU-Chest data set.

hundred-thousand to millions depending on the network size) to achieve an adequate generalization capability. Albeit, we can still leverage their power by finding properly pretrained models for similar problems. To this end, we use a state-of-the-art pneumonia detection network, CheXNet, whose details are summarized in Section III-A. With the pretrained model, we extract 1024-long vectors, right after the last average pooling layer. After data normalization (zero mean and unit variance), we obtain a feature vector $\mathbf{s} \in \mathbb{R}^{d=1024}$.

A dimensionality reduction PCA is applied to \mathbf{s} in order to get the query sample, $\mathbf{y} = \mathbf{A}\mathbf{s} \in \mathbb{R}^m$, where $\mathbf{A} \in \mathbb{R}^{m \times d}$ is PCA matrix ($m < d$).

C. Proposed CSEN-Based Classification

Considering the limited number of training data in our COVID-19 data set, a representation-based classification can be applied hereafter to obtain the class of \mathbf{y} using the dictionary Φ (in the form of $\mathbf{D} = \mathbf{A}\Phi$), whose columns are stacked training samples with class-specific locations.

As discussed earlier, SRC is an SE problem which is expected to be an easier task than an SSR problem. On the other hand, even if the exact signal recovery is not possible in noisy cases or in cases where $\hat{\mathbf{x}}$ is not exactly but approximately sparse (which is the case almost all the time in dictionary-based classification problems), it is still possible to recover the support set exactly [25], [38], [45], [46] or partially [46]–[48]. However, many works in the literature dealing with SE problems tend to first apply a sparse recovery technique on \mathbf{y} to first get $\hat{\mathbf{x}}$, then use simple thresholding over $\hat{\mathbf{x}}$ to obtain a sparse SE, $\hat{\Lambda}$. However, SSR techniques such as ℓ_1 -minimization are rather slow and their performance varies from one SRR tool to another [9]. In our previous work [9], we proposed an alternative solution for this iterative sparse recovery approach which aims to learn a direct mapping from a test sample \mathbf{y} to the corresponding support set $\hat{\Lambda}$. Along with

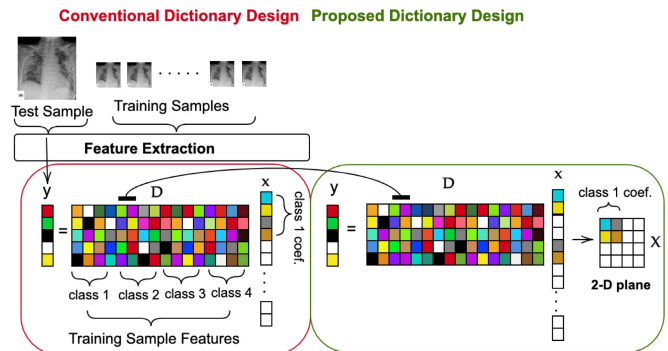


Fig. 5. Illustration of proposed dictionary design versus conventional design in representation-based classifiers.

the speed and stability compared to conventional SSR-based techniques and recent deep learning-based SSR solutions, CSEN has the crucial advantage of having a compact design that can achieve a good performance level even over scarce training data.

Mathematically speaking, an ideal CSEN is supposed to yield a binary mask $\mathbf{v} \in \{0, 1\}^n$

$$v_i = 1 \text{ if } i \in \Lambda \quad (4)$$

which indicates the true support, i.e., $\Lambda = \{i \in \{1, 2, \dots, n\} : v_i = 1\}$. In order to approximate this ideal case, a CSEN network, $\mathcal{P}(\mathbf{y}, \mathbf{D})$ produces a probability vector \mathbf{p} which returns a measure about the probability of each index being in Λ such that $p_i \in [0, 1]$. Having the estimated probability map, estimating the support can easily be done via $\hat{\Lambda} = \{i \in \{1, 2, \dots, n\} : p_i > \tau\}$, by thresholding \mathbf{p} with τ where τ is a fixed threshold.

A CSEN is composed of fully convolutional layers, and as input it takes a proxy, $\tilde{\mathbf{x}}$, of sparse coefficient vector, which is a coarse estimation of \mathbf{x} , i.e., $(\mathbf{D}^T \mathbf{D} + \lambda \mathbf{I})^{-1} \mathbf{D}^T \mathbf{y}$ or simply $\tilde{\mathbf{x}} = \mathbf{D}^T \mathbf{y}$. Then, it yields the aforementioned probability like vector \mathbf{p} via fully convolutional layers. Using such a proxy of \mathbf{x} , instead of making inference directly on \mathbf{y} has also studied in a few more recent studies. For instance, in [49] and [50], the authors proposed reconstruction-free image classification from compressively sensed images. Alternatively, one may design a network to learn proxy $\tilde{\mathbf{x}}$ by fully connected dense layers [49]. However, it increases the computational complexity and may result in an even over-fitting problem with scarce training data [9].

The input vector $\tilde{\mathbf{x}}$ is reshaped to have a 2-D plane representation in order to use it with 2-D convolutional layers. This transformation is performed via reordering the indices of the atoms in such a way that the nonzero elements of the representation vector \mathbf{x} for a specific class come together in the 2-D plane. A representative illustration of the proposed dictionary design compared to the traditional one is shown in Fig. 5.

Hereafter, the proxy $\tilde{\mathbf{x}}$ is convolved with the weight kernels, connecting the input with the next layer with N_l filters to yield the inputs of the next layer, with the biases \mathbf{b}_l as follows:

$$\mathbf{f}_l = \left\{ S_l \left(\text{ReLu}(\mathbf{b}_l^i + \mathbf{w}_l^i * \tilde{\mathbf{x}}) \right) \right\}_{i=1}^{N_l} \quad (5)$$

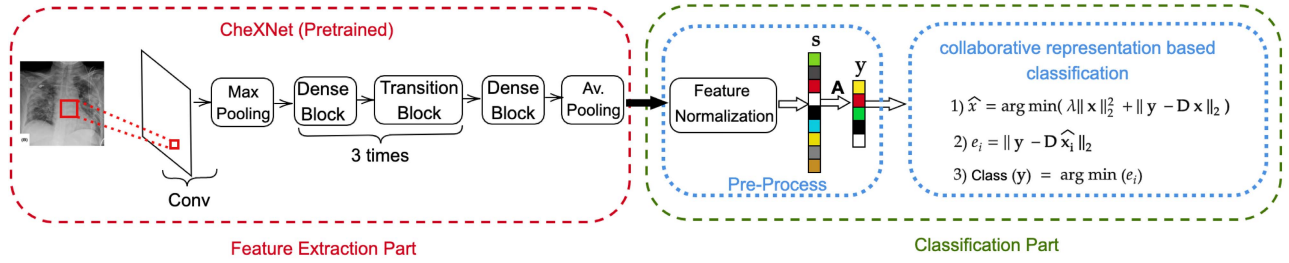


Fig. 6. Baseline Approach I: CRC is fed by deep learning-based extracted features that are preprocessed.

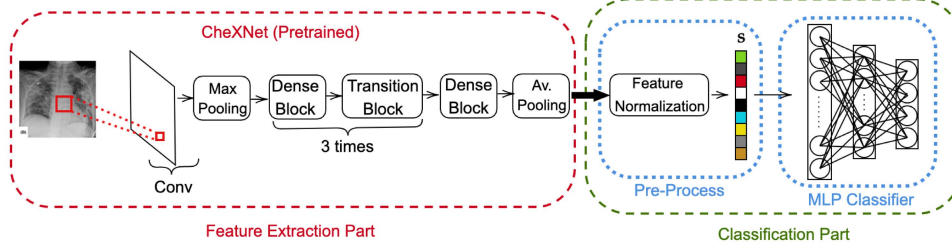


Fig. 7. Baseline Approach II: A 5-layer MLP layer is used over the features of CheXNet.

where \mathbf{b}_1 is the weight bias, $S_1(\cdot)$ is either identity or sub-sampling operator predefined according to network structure and $\text{ReLU}(x) = \max(0, x)$. For other layers, i.e., $l > 2$, the k th feature map of layer l is defined as

$$\mathbf{f}_l^k = S_l \left(\text{ReLU} \left(\mathbf{b}_l^k + \sum_i^{N_{l-1}} \mathbf{w}_l^{ik} * \mathbf{f}_{l-1}^i \right) \right) \quad (6)$$

where $S_l(\cdot)$ is either identity operator or one the operations from down- and up-sampling and N_l is the number of feature maps in l th layer. Therefore, the trainable parameters of CSEN will be: $\Theta_{\text{CSEN}} = \{ \{ \mathbf{w}_1^i, \mathbf{b}_1^i \}_{i=1}^{N_1}, \{ \mathbf{w}_2^i, \mathbf{b}_2^i \}_{i=1}^{N_2}, \dots, \{ \mathbf{w}_L^i, \mathbf{b}_L^i \}_{i=1}^{N_L} \}$ for an L layer CSEN design.

In developing the dictionary that is to be used in the SRC, the training samples are stacked-in by grouping them according to their classes. Thus, instead of using traditional ℓ_1 -minimization formulation as in (3), the following group ℓ_1 -minimization formulation may result in increased classification accuracy:

$$\min_{\mathbf{x}} \left\{ \|\mathbf{D}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \sum_{i=1}^c \|\mathbf{x}_{Gi}\|_2 \right\} \quad (7)$$

where \mathbf{x}_{Gi} is the group of coefficients from the i th class. In this manner, one possible cost function for a SE network would be

$$E(\mathbf{x}) = \sum_p (\mathcal{P}_{\Theta}(\tilde{\mathbf{x}})_p - v_p)^2 + \lambda \sum_{i=1}^c \|\mathcal{P}_{\Theta}(\tilde{\mathbf{x}})_{Gi}\|_2 \quad (8)$$

where $\mathcal{P}_{\Theta}(\tilde{\mathbf{x}})_p$ is network output at location p and v_p is the ground truth binary mask of the sparse code \mathbf{x} . Due to its high computational complexity, we approximate the cost function in (8) with a simpler average pooling layer after convolutional layer, which can produce directly the estimated class in our CSEN design. An illustration of proposed CSEN-based COVID-19 recognition is shown in Fig. 3.

D. Competing Methods

This section summarizes the competing methods that are selected among numerous alternatives due to their superior performance levels obtained in similar problems. For fair comparative evaluations, all classification methods have the same input feature vectors fed to the proposed CSENs.

1) *Collaborative Representation-Based Classification*: As a possible competing technique to the proposed CSEN-based technique which is a hybrid method, CRC [6] is a direct and representation-based classification method that can be applied to this problem as shown in Fig. 6. It is a noniterative SE technique, that satisfies faster and comparable classification performance with SRC while it is more stable compared to existing iterative sparse recovery tools as it is shown in [9]. In the first step of CRC, the tradeoff parameter of the regularized least-square solution is set as $\lambda = 2 * 10^{-12}$. In order to obtain the best possible λ , a grid search was made in the range $[10^{-15}, 10^{-1}]$ with a log scale.

2) *Multilayer Perceptron (MLP) Classification*: The proposed COVID-19 recognition pipeline can be modified by replacing CSEN or CRC part with another classifier. As one of the most-common classifiers, a 4-hidden layer multilayer perceptron (MLP) is used for this problem as shown in Fig. 7. For training, we used back-propagation (BP) with Adam optimization technique [51]. The network and training hyperparameters are as follows: learning rate, $\alpha = 10^{-4}$, and moment updates $\beta_1 = 0.9$, $\beta_2 = 0.999$, and 50 as the number of epochs. Fig. 8 illustrates the network configuration in detail. This network configuration has achieved the best performance among others (deeper and shallower) where deep configurations have suffered from *over-fitting* while the shallow ones exhibit an inferior learning performance.

3) *Support Vector Machines (SVMs)*: For a multiclass problem, the first objective is to select the SVM topology for ensemble learning: one-versus-one or one-versus-all. In order to find the optimal topology and the hyperparameters (e.g., kernel type and its parameters) we first performed a grid-search with the following variations and setting: kernel function

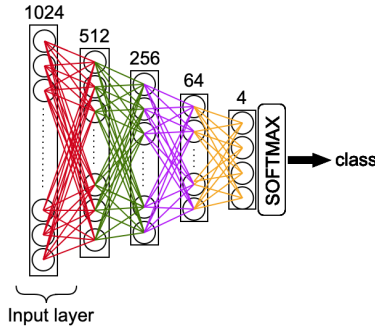


Fig. 8. MLP configuration.

TABLE I

CLASSIFICATION PERFORMANCES OF THE PROPOSED CSEN AND COMPETING METHODS. THE BEST COVID-19 RECOGNITION RATES ARE HIGHLIGHTED

		k-NN	SVM	MLP	CRC	ReconNet	CSEN1	CSEN2
Accuracy	Bacterial	0.777	0.780	0.763	0.820	0.765	0.793	0.794
	Viral	0.801	0.787	0.765	0.827	0.785	0.805	0.803
	Normal	0.903	0.934	0.933	0.928	0.918	0.926	0.927
	COVID-19	0.950	0.945	0.949	0.955	0.936	0.955	0.959
TN	Bacterial	3166	3219	3114	3063	3180	3177	3173
	Viral	4123	3965	3923	4385	4005	4109	4091
	Normal	4253	4444	4442	4380	4364	4388	4396
	COVID-19	5525	5489	5522	5554	5435	5548	5572
TP	Bacterial	1720	1687	1680	2091	1629	1810	1818
	Viral	909	979	884	816	928	954	959
	Normal	1420	1427	1421	1456	1407	1431	1428
	COVID-19	446	452	444	447	448	455	455
FP	Bacterial	360	307	412	463	346	349	353
	Viral	678	836	878	416	796	692	710
	Normal	454	263	265	327	343	319	311
	COVID-19	299	335	302	270	389	276	252
FN	Bacterial	1040	1073	1080	669	1131	950	942
	Viral	576	506	601	669	557	531	526
	Normal	159	152	158	123	172	148	151
	COVID-19	16	10	18	15	14	7	7
Specificity	Bacterial	0.623	0.611	0.609	0.758	0.590	0.656	0.659
	Viral	0.612	0.660	0.595	0.550	0.625	0.642	0.646
	Normal	0.899	0.904	0.900	0.922	0.891	0.906	0.904
	COVID-19	0.965	0.978	0.961	0.968	0.970	0.985	0.985
F1-score	Bacterial	0.898	0.913	0.883	0.869	0.902	0.901	0.900
	Viral	0.859	0.826	0.817	0.913	0.834	0.856	.852
	Normal	0.904	0.944	0.944	0.931	0.927	0.932	0.934
	COVID-19	0.949	0.943	0.948	0.954	0.933	0.953	0.957
F1-score	Bacterial	0.711	0.710	0.693	0.787	0.688	0.736	0.737
	Viral	0.592	0.593	0.545	0.601	0.578	0.609	0.608
	Normal	0.823	0.873	0.870	0.866	0.845	0.860	0.861
	COVID-19	0.740	0.724	0.735	0.758	0.690	0.763	0.778

{linear, radial basis function (RBF)}, box constraint (C parameter) in the range $[1, 10^3]$ with a log scale, and kernel scale (γ for the RBF kernel) in the range $[10^{-4}, 10^{-2}]$ with a log scale.

4) *k-Nearest-Neighbor (k-NN)*: Finally, we use a traditional approach, k -nearest neighbor (k -NN) is used with PCA dimensionality reduction. In a similar fashion, the distance metric and the k -value are optimized by a prior grid-search. The following distance metrics are evaluated: City-block, Chebyshev, correlation, cosine, Euclidean, Hamming, Jaccard, Mahalanobis, Minkowski, standardized Euclidean, and Spearman metrics. The k -value is varied within the range of $[1, 4416]$ with a log scale.

V. EXPERIMENTAL RESULTS

A. Experimental Setup

We have performed our experiments over the QaTa-Cov19 data set, which consists of normal and three pneumonia classes: bacterial, viral, and COVID-19.

TABLE II
NUMBER OF IMAGES PER CLASS AND PER-FOLD BEFORE AND AFTER DATA AUGMENTATION

Class	# of Samples	Training Samples	Augmented Training Samples	Test Samples
Bacterial Pneumonia	2760	2208	2208	552
Viral Pneumonia	1485	1188	2208	297
Normal	1579	1263	2208	316
COVID-19	462	370	2208	92
Total	6286	5029	8832	1257

The proposed approach is evaluated using a stratified fivefold cross-validation (CV) scheme with a ratio of 80% for training and 20% for the test (unseen folds) splits, respectively.

Table II shows the number of X-ray images per class in the QaTa-Cov19 data set. Since the data set is unbalanced, we have applied data augmentation to the training set in order to balance the size of each class in the train set. Therefore, the X-ray images in viral and COVID-19 pneumonia and normal classes are augmented up to the same number as the bacterial pneumonia class in the train set. We use Image Data Generator by Keras to perform data augmentation by randomly rotating the X-ray images in a range of 10° , randomly shifting images both horizontally and vertically within the interval of $[-0.1, +0.1]$. In each CV fold, we use a total of 8832 and 1257 images in the train and test (unseen in the fold) sets, respectively.

The experimental evaluations of SVM, k -NN, and CRC are performed using MATLAB version 2019a, running on PC with Intel® i7-8650U CPU and 32 GB system memory. On the other hand, MLP and CSEN methods are implemented using Tensorflow library [52] with Python on NVidia® TITAN-X GPU card. For the CSEN training, ADAM optimizer [51] is used with the proposed default learning parameters: learning rate, $\alpha = 10^{-3}$, and moment updates $\beta_1 = 0.9$, $\beta_2 = 0.999$ with only 15 back-propagation epochs. Neither grid-search nor any other parameter or configuration optimization was performed for CSEN.

B. Experimental Results

The same network configurations are used for CSEN as in [9]. Accordingly, we use two compact CSEN designs: CSEN1 and CSEN2, respectively. The first CSEN network consists of only two hidden convolutional layers, the first layer has 48 neurons and the second has 24. ReLU activation function is used in the hidden layers and the filter size was 3×3 . On the other hand, CSEN2 uses max-pooling and has one additional hidden layer with 24 neurons to perform transposed-convolution. CSEN1 and CSEN2 are compared against the 6 competing methods under the same experimental setup.

For the dictionary construction in Φ each CSEN design, 625 images for each class (from the augmented training samples per fold) are stacked in such way that the representation coefficient in the 2-D plane, \mathbf{X} has 50×50 size as shown in Fig. 5. The rest of the images in the training set are used to train each CSEN, i.e., 1583 samples from each class. We use PCA dimensional reduction matrix, \mathbf{A} with the compression ratio, $CR = (m/d) = 0.5$. Therefore, we have 512×2500 equivalent

TABLE III
NUMBER OF NETWORK PARAMETERS OF EACH METHOD

	MLP	CSEN1	CSEN2	ReconNet
# of trainable parameters	672,836	11,089	16,297	22,914

dictionary, \mathbf{D} , and 2500×512 denoiser $\mathbf{B} = (\mathbf{D}^T \mathbf{D} + \lambda \mathbf{I})^{-1} \mathbf{D}^T$ to obtain a coarse estimation of the representation (sparse in the ideal case) coefficients, $\tilde{\mathbf{x}} \in \mathbb{R}^{n=2500}$. Hereafter, the CSEN networks are trained to obtain the class information of \mathbf{y} from input $\tilde{\mathbf{x}}$ as illustrated in Fig. 3.

Due to the lack of other learning-based SE studies in the literature, we chose a deeper network compared to CSEN designs to investigate the role of network depth in this problem. ReconNet [53] was proposed as a noniterative deep learning solution to CS problem, i.e., $\hat{\mathbf{s}} \leftarrow \mathcal{P}(\mathbf{y})$ and it is one of the state of the art in compressively sensed image recognition task. It consists of six fully convolutional layers and one dense layer in front of the convolutional ones, which act as the learned denoiser for the mapping from $\mathbf{y} \in \mathbb{R}^m$ to $\tilde{\mathbf{s}} \in \mathbb{R}^d$. Then, the convolutional layers are responsible for producing the reconstructed signal, $\hat{\mathbf{s}}$ from $\tilde{\mathbf{s}}$. Therefore, by replacing this dense layer with the denoiser matrix \mathbf{B} , this network can be used as a competing method.

Both CSEN and the modified ReconNet use $\tilde{\mathbf{x}}$ as an input, which is produced using an equivalent dictionary \mathbf{D} and its pseudo-inverse matrix \mathbf{B} .

In designing the dictionary of the CRC system, all training samples are stacked in the dictionary, Φ , i.e., 2208 samples from each class. The same PCA matrix used in CSEN-based recognition, \mathbf{A} is applied to features, $\mathbf{s} \in \mathbb{R}^{d=1024}$. Therefore, a dictionary \mathbf{D} of size 512×8832 and the corresponding denoiser matrix \mathbf{B} of size 8832×512 are used in the CRC framework.

Overall, the confusion matrix elements are formed as follows: true positive (TP): the number of correctly detected positive class members, true negative (TN): the number of correctly detected negative class samples, false positive (FP): the number of misclassified negative class members as positive, and FN: the number of misclassified positive class samples as negative (i.e., missed positive cases). Then, the standard performance evaluation metrics are defined as follows:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (9)$$

where sensitivity (or Recall) is the rate of correctly detected positive samples in the positive class

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (10)$$

where specificity is the ratio of accurately detected negative class samples to all negative class

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (11)$$

where precision is the rate of correctly classified positive class samples among all the members classified as a positive sample

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TN} + \text{TP} + \text{FP} + \text{FN}} \quad (12)$$

TABLE IV
COMPUTATION TIMES (SEC) OF EACH METHOD OVER 1257 TEST IMAGES

	CRC (light)	CRC	CSEN1	CSEN2	ReconNet	MLP
Computation Time (in sec.)	13.4176	40.7878	0.2196	0.2272	0.2993	0.2935

TABLE V
OVERALL (CUMULATIVE) CONFUSION MATRIX OF THE PROPOSED RECOGNITION SCHEME

CSEN2	Predicted				
		Bacterial	Viral	Normal	COVID-19
Real	Bacterial	1818	636	180	126
	Viral	338	959	127	61
	Normal	15	71	1428	65
	COVID-19	0	3	4	455

where accuracy is the ratio of correctly classified elements among all the data

$$F(\beta) = (1 + \beta^2) \frac{(\text{Precision} + \text{Sensitivity})}{(\beta^2 * \text{Precision}) + \text{Sensitivity}} \quad (13)$$

where F -score is defined by the weighting parameter β . The $F1$ -score is calculated with $\beta = 1$, which is the harmonic average of precision and sensitivity.

The classification performance of the proposed CSEN-based approach and the competing methods is presented in Table I. As can be easily observed from Table I, the proposed approaches surpass all competing methods in COVID-19 recognition performance by achieving 98.5% sensitivity, and over 95% specificity. As shown in Table III, compared to MLP and ReconNet, the proposed CSEN designs are very compact and computationally efficient. This is evident in Table IV where the computational complexity (measured as total computation, time over the 1257 test images) is reported.

Finally, Table V presents the overall (cumulative) confusion matrix of the proposed CSEN-based COVID-19 recognition approach over the new QaTa-Cov19 data set. The most critical misclassifications are the false-positives, i.e., the misclassified COVID-19 X-ray images. The confusion matrix shows that the proposed approach has misclassified seven COVID-19 images (out of 462). The 3 out of 7 misclassifications are still in “viral pneumonia” category, which can be an expected confusion due to the viral nature of COVID-19. However, the other four cases are misclassified as “Normal” which is indeed a severe clinical misdiagnosis. A close look at these false-negatives in Fig. 9 reveals the fact that they are indeed very similar to normal images where typical COVID-19 patterns are hardly visible even by an expert’s naked eye. It is possible that these images come from patients who were in the very early stages of COVID-19.

VI. DISCUSSION

A. CRC Versus CSEN

When compared against CRC in particular, CSEN-based classification has two advantages; computational efficiency and, a superior COVID-19 recognition performance. The computational efficiency comes from the fact that a larger size dictionary matrix (of the size of 512×8832) is used

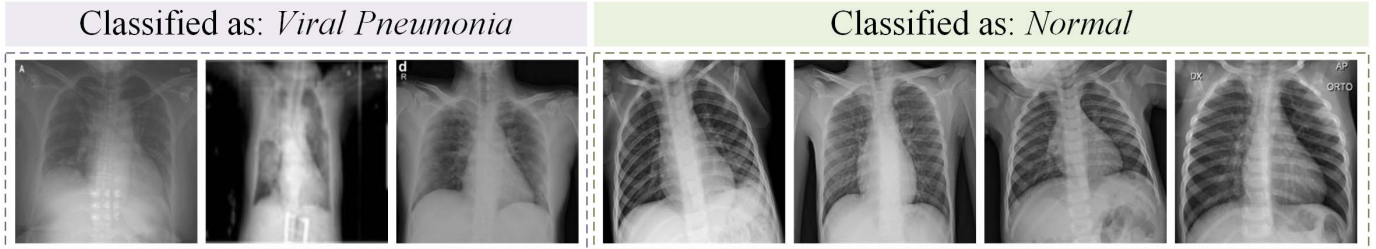


Fig. 9. FNs of the proposed COVID-19 recognition scheme.

TABLE VI

PERFORMANCE OF CRC ALGORITHM WHEN THE DICTIONARY (SIZE OF 625 PER CLASS) THAT IS USED IN CSEN IS USED

	CRC (Light)		
	Accuracy	Sensitivity	Specificity
Bacterial	0.8129	0.7464	0.8650
Viral	0.8163	0.5461	0.8998
Normal	0.9267	0.9170	0.9299
COVID-19	0.9564	0.9394	0.9578

in CRC and hence, this requires more computations in terms of matrix-vector multiplications. Furthermore, saving the trainable parameters ($\sim 16k$) and a light dictionary matrix coefficients ($\sim 1280k$) in the test device is more memory efficient compared to saving coefficients ($\sim 4521k$) of larger size dictionary used in CRC.

For further analysis, we also tested the CRC framework by using the light dictionary (of size 512×2500) used in CSEN-based recognition. We called it CRC (light), and as it can be seen in Table VI, the performance of CRC further reduced, and there was no significant improvement concerning the computational cost. When it comes to creating deeper convolutional layers instead of using CSEN designs, such as the modified ReconNet, the results presented in Table I shows us that compact CSEN structures are indeed preferable to achieve superior classification performances compared to deeper networks.

B. Compact Versus Deep CSENs

Representation-based classifications are known for providing satisfactory performance when it comes to limited size data sets. On the other hand, deep artificial NNs usually require a large training set to achieve a satisfactory generalization capability.

In a representation-based (dictionary) classification scheme when the dictionary size getting bigger (increase the number of training samples), the computational complexity of the method drastically increases. The proposed CSEN is an alternative approach to handle both moderate and scarce data sets via compact as possible NN structures for the dictionary-based classification.

Since there is no other learning-based SE method except CSEN in the literature, we chose ReconNet as a possible competing algorithm for this problem as explained in detail in Section V. ReconNet has six fully convolution layers. As an ablation study, we also add more hidden layers to proposed CSEN models to compare: CSEN3 and CSEN4 models were obtained by adding one and two hidden layers to CSEN2, respectively, after the transposed convolutional layer.

TABLE VII

PERFORMANCE OF ALTERNATIVE DEEPER DESIGNS COMPARED TO COMPACT CSENS

	Accuracy		Sensitivity		Specificity	
	CSEN3	CSEN4	CSEN3	CSEN4	CSEN3	CSEN4
Bacterial	0.793	0.792	0.651	0.653	0.904	0.900
Viral	0.808	0.805	0.642	0.638	0.859	0.856
Normal	0.922	0.921	0.907	0.899	0.927	0.928
Covid-19	0.954	0.954	0.990	0.987	0.951	0.952

TABLE VIII

NUMBER OF NETWORK PARAMETERS OF COMPETING SE NETWORKS

	CSEN1	CSEN2	CSEN3	ReconNet	CSEN 4
# of trainable parameters	11,089	16,297	21,505	22,914	26,713

Additional layers have 24 neurons, ReLu activation functions and filter size 3×3 . As we can observe from Tables VII and VIII, the proposed compact designs, CSEN1 and CSEN2, both surpass deeper counterparts both in performance and the required number of parameters.

VII. CONCLUSION

The commonly used methods in COVID-19 diagnosis, namely RT-PCR and CT have certain limitations and drawbacks such as long processing times and unacceptably high misdiagnosis rates. These drawbacks are also shared by most of the recent works in the literature based on deep learning due to data scarcity from the COVID-19 cases. Although deep learning-based recognition techniques are dominant in computer vision where they achieved state-of-the-art performance, their performance degrades fast due to data scarcity, which is the reality in this problem at hand. This study aims to address such limitations by proposing a robust and highly accurate COVID-19 recognition approach directly from X-ray images. The proposed approach is based on the CSEN that can be seen as a bridge between deep learning models and representation-based methods. CSEN uses both a dictionary and a set of training samples to learn a direct mapping from the query samples to the sparse support set of representation coefficients. With this unique ability and having the advantage of a compact network, the proposed CSEN-based COVID-19 recognition systems surpass the competing methods and achieve over 98% sensitivity and over 95% specificity. Furthermore, they yield the most computationally efficient scheme in terms of speed and memory.

ACKNOWLEDGMENT

The authors would like to thank the following medical doctor team for their generous feedbacks and continuous

proof reading: Khalid Hameed is a MD in Reem Medical Center, Doha, Qatar. Tahir Hamid is consultant cardiologist in Hamad Medical Corporation Hospital and with Weill Cornell Medicine-Qatar, Doha. Rashid Mazhar is a MD in Hamad Medical Corporation Hospital, Doha, Qatar.

REFERENCES

- [1] L. Pellis *et al.*, “Challenges in control of COVID-19: Short doubling time and long delay to effect of interventions,” 2020, *arXiv:2004.00117*. [Online]. Available: <http://arxiv.org/abs/2004.00117>
- [2] F. Zhou *et al.*, “Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: A retrospective cohort study,” *Lancet*, vol. 395, no. 10229, pp. 1054–1062, Mar. 2020.
- [3] Y. Fang *et al.*, “Sensitivity of chest CT for COVID-19: Comparison to RT-PCR,” *Radiology*, vol. 296, no. 2, pp. E115–E117, Aug. 2020.
- [4] K. A. Erickson, K. Mackenzie, and A. Marshall, “Advanced but expensive technology. Balancing affordability with access in rural areas,” *Can. Family Physician Medecin de Famille Canadien*, vol. 39, pp. 28–30, Jan. 1993.
- [5] World Health Organization, “Laboratory testing for coronavirus disease (COVID-19) in suspected human cases: Interim guidance,” World Health Org., Tech. Rep. WHO/COVID-19/laboratory/2020.5, Mar. 2020.
- [6] L. Zhang, M. Yang, and X. Feng, “Sparse representation or collaborative representation: Which helps face recognition?” in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 471–478.
- [7] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, “Robust face recognition via sparse representation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.
- [8] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, “Sparse representation for computer vision and pattern recognition,” *Proc. IEEE*, vol. 98, no. 6, pp. 1031–1044, Jun. 2010.
- [9] M. Yamac, M. Ahishali, S. Kiranyaz, and M. Gabbouj, “Convolutional sparse support estimator network (CSEN) from energy efficient support estimation to learning-aided compressive sensing,” 2020, *arXiv:2003.00768*. [Online]. Available: <http://arxiv.org/abs/2003.00768>
- [10] B. de Fourier and J. B. Joseph, *Théorie Analytique de la Chaleur*. Firmin Didot, 1822.
- [11] S. G. Mallat and Z. Zhang, “Matching pursuits with time-frequency dictionaries,” *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.
- [12] J.-L. Starck, E. J. Candes, and D. L. Donoho, “The curvelet transform for image denoising,” *IEEE Trans. Image Process.*, vol. 11, no. 6, pp. 670–684, Jun. 2002.
- [13] J. Yang, K. Yu, Y. Gong, and T. Huang, “Linear spatial pyramid matching using sparse coding for image classification,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1794–1801.
- [14] A. Adler, M. Elad, Y. Hel-Or, and E. Rivlin, “Sparse coding with anomaly detection,” *J. Signal Process. Syst.*, vol. 79, no. 2, pp. 179–188, May 2015.
- [15] D. Carrera, G. Boracchi, A. Foi, and B. Wohlberg, “Detecting anomalous structures by convolutional sparse models,” in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2015, pp. 1–8.
- [16] W. Wen, C. Wu, Y. Wang, Y. Chen, and H. Li, “Learning structured sparsity in deep neural networks,” in *Adv. neural Inf. Process. Syst.*, 2016, pp. 2074–2082.
- [17] D. L. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [18] E. J. Candès, “Compressive sampling,” in *Proc. Int. Congr. Math.*, vol. 3. Madrid, Spain, 2006, pp. 1433–1452.
- [19] D. L. Donoho and M. Elad, “Optimally sparse representation in general (nonorthogonal) dictionaries via l_1 minimization,” *Proc. Nat. Acad. Sci. USA*, vol. 100, no. 5, pp. 2197–2202, Mar. 2003.
- [20] A. Cohen, W. Dahmen, and R. DeVore, “Compressed sensing and best K -term approximation,” *J. Amer. Math. Soc.*, vol. 22, no. 1, pp. 211–231, 2009.
- [21] H. Rauhut, “Compressive sensing and structured random matrices,” in *Theoretical Foundations and Numerical Methods for Sparse Recovery* (Radon Series on Computational and Applied Mathematics), vol. 9, M. Fornasier, Ed. deGruyter, 2010, pp. 1–92.
- [22] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, Jan. 2001.
- [23] E. J. Candès and T. Tao, “Decoding by linear programming,” *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
- [24] E. J. Candès, “The restricted isometry property and its implications for compressed sensing,” *Comp. Rendus Mathématique*, vol. 346, nos. 9–10, pp. 589–592, May 2008.
- [25] W. Wang, M. J. Wainwright, and K. Ramchandran, “Information-theoretic limits on sparse support recovery: Dense versus sparse measurements,” in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2008, pp. 2197–2201.
- [26] J. Haupt and R. Baraniuk, “Robust support recovery using sparse compressive sensing matrices,” in *Proc. 45th Annu. Conf. Inf. Sci. Syst.*, Mar. 2011, pp. 1–6.
- [27] P. Rajpurkar *et al.*, “CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning,” 2017, *arXiv:1711.05225*. [Online]. Available: <http://arxiv.org/abs/1711.05225>
- [28] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.
- [29] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A large-scale hierarchical image database,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
- [30] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, “ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2097–2106.
- [31] S. Shekhar, V. M. Patel, N. M. Nasrabadi, and R. Chellappa, “Joint sparse representation for robust multimodal biometrics recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 1, pp. 113–126, Jan. 2014.
- [32] X. Mei and H. Ling, “Robust visual tracking and vehicle classification via sparse representation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 11, pp. 2259–2272, Nov. 2011.
- [33] T. Guha and R. K. Ward, “Learning sparse representations for human action recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 8, pp. 1576–1588, Aug. 2012.
- [34] W. Li and Q. Du, “A survey on representation-based classification and detection in hyperspectral remote sensing imagery,” *Pattern Recognit. Lett.*, vol. 83, pp. 115–123, Nov. 2016.
- [35] M. E. H. Chowdhury *et al.*, “Can AI help in screening viral and COVID-19 pneumonia?” 2020, *arXiv:2003.13145*. [Online]. Available: <http://arxiv.org/abs/2003.13145>
- [36] I. D. Apostolopoulos and T. A. Mpesiana, “Covid-19: Automatic detection from X-ray images utilizing transfer learning with convolutional neural networks,” *Phys. Eng. Sci. Med.*, vol. 43, no. 2, pp. 635–640, Jun. 2020.
- [37] L. O. Hall, R. Paul, D. B. Goldgof, and G. M. Goldgof, “Finding covid-19 from chest X-rays using deep learning on a small dataset,” 2020, *arXiv:2004.02060*. [Online]. Available: <http://arxiv.org/abs/2004.02060>
- [38] M. Wainwright, “Information-theoretic bounds on sparsity recovery in the high-dimensional and noisy setting,” in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2007, pp. 961–965.
- [39] J. P. Cohen, P. Morrison, and L. Dao, “COVID-19 image data collection,” 2020, *arXiv:2003.11597*. [Online]. Available: <http://arxiv.org/abs/2003.11597>
- [40] (2020). *COVID-19 database*. [Online]. Available: <https://www.sirm.org/category/senza-categoria/covid-19/>
- [41] (2020). [Online]. Available: <https://radiopaedia.org/playlists/25975?>
- [42] (2020). [Online]. Available: <https://threadreaderapp.com/thread/1243928581983670272.html>
- [43] J. C. Montreal. (2020). *COVID-Chestxray Database*, [Online]. Available: <https://github.com/iecc8023/covid-chestxray-dataset>
- [44] P. Mooney. (2018). *Chest X-ray Images (Pneumonia)*. kaggle, Marzo. [Online]. Available: <https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>
- [45] K. Rahnema Rad, “Nearly sharp sufficient conditions on exact sparsity pattern recovery,” *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4672–4679, Jul. 2011.
- [46] J. Scarlett and V. Cevher, “Limits on support recovery with probabilistic models: An information-theoretic framework,” *IEEE Trans. Inf. Theory*, vol. 63, no. 1, pp. 593–620, Jan. 2017.
- [47] G. Reeves and M. Gastpar, “Sampling bounds for sparse support recovery in the presence of noise,” in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2008, pp. 2187–2191.
- [48] G. Reeves and M. C. Gastpar, “Approximate sparsity pattern recovery: Information-theoretic lower bounds,” *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3451–3465, Jun. 2013.

- [49] A. Degerli, S. Aslan, M. Yamac, B. Sankur, and M. Gabbouj, "Compressively sensed image recognition," in *Proc. 7th Eur. Workshop Vis. Inf. Process. (EUVIP)*, Nov. 2018, pp. 1–6.
- [50] S. Lohit, K. Kulkarni, and P. Turaga, "Direct inference on compressive measurements using convolutional neural networks," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 1913–1917.
- [51] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: <http://arxiv.org/abs/1412.6980>
- [52] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," 2016, *arXiv:1603.04467*. [Online]. Available: <http://arxiv.org/abs/1603.04467>
- [53] K. Kulkarni, S. Lohit, P. Turaga, R. Kerviche, and A. Ashok, "ReconNet: Non-iterative reconstruction of images from compressively sensed measurements," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 449–458.



Mehmet Yamac received the B.S. degree in electrical and electronics engineering from Anadolu University, Eskisehir, Turkey, in 2009, and the M.S. degree in electrical and electronics engineering from Bogazici University, Istanbul, Turkey, in 2014. He is currently pursuing the Ph.D. degree with the Department of Computing Sciences, Tampere University, Tampere, Finland.

He was a Research and Teaching Assistant with Bogazici University from 2012 to 2017 and a Researcher with Tampere University from 2017 to

2020. He is currently working as a Senior Researcher with Huawei Technologies Oy, Helsinki, Finland. He has coauthored the articles nominated for the "Best Paper Award" or the "Student Best Paper Award" in EUVIP 2018 and EUSIPCO 2019. His research interests are computer and machine vision, machine learning, and compressive sensing.



Mete Ahishali received the B.Sc. degree (Hons.) in electrical and electronics engineering from the Smyrna University of Economics, Smyrna, Turkey, in 2017, and the M.Sc. degree (Hons.) in data engineering and machine learning from Tampere University, Tampere, Finland, in 2019, where he is currently pursuing the Ph.D. degree in computing and electrical engineering.

Since 2017, he has been working as a Researcher with the Signal Analysis and Machine Intelligence Research Group under the supervision of

Prof. Gabbouj. His research interests are pattern recognition, machine learning, and semantic segmentation with applications in computer vision, remote sensing, and biomedical images.



Aysen Degerli received the B.Sc. degree (Hons.) in electrical and electronics engineering from the Smyrna University of Economics, Smyrna, Turkey, in 2017, and the M.Sc. degree (Hons.) in data engineering and machine learning from Tampere University, Tampere, Finland, in 2019, where she is currently pursuing the Ph.D. degree in computing and electrical engineering with the Signal Analysis and Machine Intelligence Research Group led by Prof. M. Gabbouj.

Her research interests include machine learning, compressive sensing, and biomedical image processing.



Serkan Kiranyaz (Senior Member, IEEE) is a Professor with Qatar University, Doha, Qatar. He published two books, five book chapters, more than 80 journal articles in high impact journals, and 100 articles in international conferences. He made contributions on evolutionary optimization, machine learning, bio-signal analysis, computer vision with applications to recognition, classification, and signal processing. He has coauthored the articles which have nominated or received the "Best Paper Award" in ICIP 2013, ICPR 2014, ICIP 2015, and IEEE

TRANSACTIONS ON SIGNAL PROCESSING (TSP) 2018. He had the most-popular articles in the years 2010 and 2016, and most-cited article in 2018 in IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING. From 2010 to 2015, he authored the 4th most-cited article of the *Neural Networks* journal. His research team has won the second and first places in PhysioNet Grand Challenges 2016 and 2017, among 48 and 75 international teams, respectively. His theoretical contributions to advance the current state of the art in modeling and representation, targeting high long-term impact, while algorithmic, system level design and implementation issues target medium and long-term challenges for the next five to ten years. He in particular aims at investigating scientific questions and inventing cutting edge solutions in "personalized biomedicine" which is in one of the most dynamic areas where science combines with technology to produce efficient signal and information processing systems.

Prof. Kiranyaz received the "Research Excellence Award" and the "Merit Award" of Qatar University in 2019.



Muhammad E. H. Chowdhury (Senior Member, IEEE) received the Ph.D. degree from the University of Nottingham, Nottingham, U.K., in 2014.

He worked as a Postdoctoral Research Fellow with the Sir Peter Mansfield Imaging Center, University of Nottingham. He is currently working as an Assistant Professor with the Department of Electrical Engineering, Qatar University, Doha, Qatar. He has two patents and published around 80 peer-reviewed journal articles, conference papers, and four book chapters. His current research interests

include biomedical instrumentation, signal processing, wearable sensors, medical image analysis, machine learning, embedded system design, and simultaneous EEG/fMRI. He is also running several QNRF grants and internal grants from Qatar University along with academic and government projects along with different national and international projects. He has worked as a Consultant for the projects entitled, "Driver Distraction Management Using Sensor Data Cloud (2013–14)," Information Society Innovation Fund (ISIF Asia).

Dr. Chowdhury received the ISIF Asia Community Choice Award 2013 for a project entitled, "Design and Development of Precision Agriculture Information System for Bangladesh." He has recently won the COVID-19 Data Set Award for his contribution to the fight against COVID-19. He is serving as an Associate Editor for IEEE ACCESS and a Topic Editor for *Frontiers in Neuroscience*.



Moncef Gabbouj (Fellow, IEEE) received the B.S. degree from Oklahoma State University, Stillwater, OK, USA, in 1985, and the M.S. and Ph.D. degrees from Purdue University, in 1986 and 1989, respectively, all in electrical engineering.

He is a Professor of signal processing with the Department of Computing Sciences, Tampere University, Tampere, Finland. He was an Academy of Finland Professor from 2011 to 2015. His research interests include big data analytics, multimedia content-based analysis, indexing and retrieval, artificial intelligence, machine learning, pattern recognition, nonlinear signal and image processing and analysis, voice conversion, and video processing and coding.

Dr. Gabbouj is a member of the Academia Europaea and the Finnish Academy of Science and Letters. He is the past Chairman of the IEEE CAS TC on DSP and the Committee Member of the IEEE Fourier Award for Signal Processing. He served as an Associate Editor and the Guest Editor of many IEEE, and international journals and a Distinguished Lecturer for the IEEE CASS. He is the Finland Site Director of the NSF IUCRC funded Center for Visual and Decision Informatics (CVDI) and leads the Artificial Intelligence Research Task Force of the Ministry of Economic Affairs and Employment funded Research Alliance on Autonomous Systems (RAAS).