MULTI-DEEP: A novel CAD system for coronavirus (COVID-19) diagnosis from CT images using multiple convolution neural networks

Omneya Attallah[1,*], Dina A. Ragab[1,2*], and Maha A. Sharkas[1]

[1]Electronics and Communications Engineering Department, Arab Academy for Science, Technology, and Maritime Transport (AASTMT), Alexandria, PO1029, Egypt.

[2]Electronic & Electrical Engineering Department, University of Strathclyde, Glasgow, G1 1XW, United Kingdom.

Corresponding author[*]: Omneya Attallah and Dina A. Ragab

Emails: o.attallah@aast.edu, dinaragab@aast.edu

Abstract

Coronavirus (COVID-19) was first observed in Wuhan, China, and quickly propagated worldwide. It is considered the supreme crisis of the present era and one of the most crucial hazards threatening worldwide health. Therefore, the early detection of COVID-19 is essential. The common way to detect COVID-19 is the reverse transcription-polymerase chain reaction (RT-PCR) test, although it has several drawbacks. Computed tomography (CT) scans can enable the early detection of suspected patients, however, the overlap between patterns of COVID-19 and other types of pneumonia makes it difficult for radiologists to diagnose COVID-19 accurately. On the other hand, deep learning (DL) techniques and especially the convolutional neural network (CNN) can classify COVID-19 and non-COVID-19 cases. In addition, DL techniques that use CT images can deliver an accurate diagnosis faster than the RT-PCR test, which consequently saves time for disease control and provides an efficient computer-aided diagnosis (CAD) system. The shortage of publicly available datasets of CT images, makes the CAD system's design a challenging task. The CAD systems in the literature are based on either individual CNN or two-fused CNNs; one used for segmentation and the other for classification and diagnosis. In this paper, a novel CAD system is proposed for diagnosing COVID-19 based on the fusion of multiple CNNs. First, an end-to-end classification is performed. Afterward, the deep features are extracted from each network individually and classified using a support vector machine (SVM) classifier. Next, principal component analysis (PCA) is applied to each deep feature set, extracted from each network. Such feature sets are then used to train an SVM classifier individually. Afterward, a selected number of principal components from each deep feature set are fused and compared with the fusion of the deep features extracted from each CNN. The results show that the proposed system is effective and capable of detecting COVID-19 and distinguishing it from non-COVID-19 cases with an accuracy of 94.7%, AUC of 0.98 (98%), sensitivity 95.6%, and specificity of 93.7%. Moreover, the results show that the system is efficient, as fusing a selected number of principal components has reduced the computational cost of the final model by almost 32%.

Keywords: Computed tomography (CT), computer-aided diagnosis (CAD), convolution neural networks (CNN), coronavirus (COVID-19), deep learning (DL).

1. Introduction

Late December 2019, the coronavirus disease (COVID-19) was detected in Wuhan City, Hubei Province, China. It then propagated worldwide, and was declared a pandemic by the World Health Organization (WHO) at the end of January 2020 (Butt et al., 2020). The pandemic of the novel coronavirus is still progressively increasing and propagating around the world. So far, the pandemic has spread to more than 30 provinces of China and other countries (Li et al., 2020b). The number of patients that were verified to have COVID-19 has increased to over 775,000 in more than 160 countries, but the number of persons diseased is likely to be much greater. Over 36,000 individuals have died from the COVID-19 disease (up to 30 March 2020) (Dong, Du & Gardner, 2020).

Nowadays, COVID-19 is considered the most significant and crucial hazard to worldwide health (Dong, Du & Gardner, 2020). The WHO encouraged efforts to reduce the spread of the infection,

52 however, many countries have faced some major care crises (Arabi, Murthy & Webb, 2020),
53 (Grasselli, Pesenti & Cecconi, 2020). The epidemic of COVID-19 led to a tremendous increase in
54 the need for hospital beds, as well as necessary medical equipment. Moreover, doctors and nurses
55 are at a high risk of infection. The new coronavirus has led to severe health problems such as heart
56 complications, acute respiratory condition, and secondary infections, in quite a great portion of
57 infected people. It consequently led to significant deaths. Therefore, the early detection and
58 initiation of treatment in severe cases is essential (Butt et al., 2020).

59 The reverse transcription-polymerase chain reaction (RT-PCR) test is a widely used tool to
60 diagnose COVID-19 cases. However, it suffers from some drawbacks, such as its time-consuming
61 procedure and expensive cost. Moreover, the sensitivity of the RT-PCR test is inadequate, resulting
62 in false negatives, and subsequently more infections (Ai et al., 2020). Furthermore, the
63 insufficient amount of tests and the need for well-equipped laboratories for analysis, could
64 critically delay the accurate detection of COVID-19 cases. This shows the unexpected challenges
65 faced, in trying to avert the spread of the virus globally (Xie et al., 2020). These limitations have
66 encouraged researches to look for a faster alternative solution to detect coronavirus.

67 Deep learning (DL) techniques are considered the new class of machine learning (ML) approaches.
68 The explanation of different DL techniques will be discussed later in the methodology section.
69 Recently, DL has been extensively used in the medical area (Zemouri, Zerhouni & Racoceanu,
70 2019). This is due to the advantage they offer over classical ML approaches. DL methods have the
71 ability to present   optimum representations and significant information from the raw images
72 without image enhancement, segmentation, and feature extraction processes. This leads to an
73 improved diagnosis process and lower complexity, when compared to classical ML approaches
74 (Attallah, Sharkas & Gadelkarim, 2020),(Fujita, 2020). Consequently, DL based-CAD systems
75 that use medical images are recommended as another tool in the diagnosis and control of the novel
76 coronavirus. Such CAD systems offer a quick and easy solution for medical diagnosis of COVID-
77 19 (Iwasawa et al., 2020). Moreover, these CAD systems have the potential to deliver an accurate
78 and fast second opinion, to help radiologists in giving an accurate diagnosis based on medical
79 images. (Ardakani et al., 2020). Furthermore, a CAD system can avoid diagnostic errors caused
80 by human which might occur due to the exertion done during clinical examinations and
81 radiologists' visual fatigue (Ragab, Sharkas & Attallah, 2019),(Ragab et al., 2019). This can assist
82 in managing the current pandemic, speed up the detection of the disease, avoid its fast spread, and
83 help doctors to improve the quality of patient management, even in extraordinary workload
84 situations.

85 Computed tomography (CT) images are well-known imaging technique that can be employed by
86 CAD systems for the early detection of the COVID-19 disease, and distinguishing it from non-
87 COVID-19 cases (Butt et al., 2020). The authors in (Fang et al., 2020) have shown that the
88 sensitivity of detecting COVID-19 using CT medical images is much higher than the RT-PCR
89 test. Several studies have demonstrated that CT is an efficient tool in detecting, visualizing, and
90 diagnosing COVID 19-disease (Fang et al., 2020),(Xie et al., 2020). The representative appearance
91 of COVID-19 in CT images allows such images to differentiate COVID-19 from non-COVID-19
92 pneumonia. This task is difficult when it depends only on human examination. This is because of
93 the overlap of specific patterns with other types of pneumonia (Shi, Han & Zheng, 2020).
94 However, DL techniques can distinguish between COVID-19 and non-COVID-19 cases (Butt et

95 al., 2020). Therefore, DL methods could be used to enhance the radiologist's diagnostic
96 capabilities for detecting COVID-19 from CT images.

97 The authors in (Butt et al., 2020) proposed the use of ResNet-23 and ResNet-18 convolutional
98 neural network (CNN) architectures. They differentiated COVID-19 and non-COVID-19 cases
99 achieving an area under the receiver-operating curve (AUC) of 99.6%, a sensitivity of 98.2%, and
100 a specificity of 92.2%. The authors in (Li et al., 2020a) introduced a method based on ResNet-50
101 to detect COVID-19 from CT images. The sensitivity, specificity, and AUC produced were 87%,
102 92%, and 0.95 (95%), respectively. EfficientNet-B4 was used in (Bai et al., 2020) to distinguish
103 between corona and non-corona virus achieving 87%, 89%, 86%, and 0.9 (90%) for accuracy,
104 sensitivity, specificity, and AUC, respectively. On the other hand, the authors in (Amyar,
105 Modzelewski & Ruan, 2020) constructed two U-Nets. The highest accuracy, sensitivity,
106 specificity, and AUC attained were 86%, 94%, 79%, and 0.93 (93%), respectively. The authors in
107 (Chen et al., 2020) employed the U-Net architecture to distinguish between COVID-19 and non-
108 COVID-19 cases. A sensitivity of 100%, a specificity of 93.6%, and an accuracy of 95.2% were
109 obtained. A U-Net was used for lung segmentation in (Zheng et al., 2020) followed by a 3D CNN
110 for predicting the probability of COVID-19. The proposed method reached a sensitivity of 90.7%,
111 a specificity of 91.1%, and an AUC of 0.959 (95.9%). Whereas, a 2D Deeplab-v1 model was used
112 in (Jin et al., 2020a) for segmenting the lung followed by the ResNet-152 model for identifying
113 COVID-19 cases. The results achieved a sensitivity of 94.1%, a specificity of 95.5%, and an AUC
114 of 0.979 (97.9%). The authors in (Jin et al., 2020b) proposed the fusion of UNet++ for segmenting
115 lesions and a ResNet-50 for classification. The sensitivity and specificity of ResNet-50 were 97.4%
116 and 92.2%, respectively. Moreover, ResNet-50 CNN was also employed in (Song et al., 2020),
117 achieving an accuracy of 86%, a sensitivity of 96%, a precision of 79%, and an AUC of 0.95
118 (95%). The authors in (Wang et al., 2020) extracted the deep features using Inception CNN
119 architecture and then used these features to train Adaboosted decision tree classifiers. An accuracy
120 of 82.9%, a sensitivity of 81%, a specificity of 84%, and an AUC of 0.9 (90%) were attained.

121 A summary of recent similar related studies is shown in Table 1. As it is clear some of the related
122 work is based on either an individual CNN or fusing two CNNs, where the first CNN is for
123 segmenting the lung and the other is for the classification and diagnosis of COVID-19. The effect
124 of fusing multiple CNNs for classification and diagnosis was not considered in the previous
125 studies. Moreover, reducing the huge deep feature space was not examined in the previous studies.
126 Therefore, in this paper, a novel CAD system based on the fusion of multiple CNN for detecting
127 COVID-19 and differentiating it from non-COVID-19 cases is proposed. The proposed CAD
128 employs four types of CNNs different from those used in the literature. These CNNs include
129 AlexNet, GoogleNet, ResNet-18, and Shuffle-Net. The proposed CAD examines the use of
130 principal component analysis (PCA), to reduce the dimensionality of the deep feature space
131 extracted from each CNN individually, and use them to train support a vector machine (SVM)
132 classifier. Then, the effect of fusing these principal components on the classification performance
133 of the SVM classifier is evaluated. The proposed CAD system classifies COVID-19 and non-
134 COVID-19 by four different scenarios. The first one is an end-to-end CNN using four fine-tuned
135 pre-trained CNNs. In the second scenario, the deep features of each CNN architecture is extracted
136 and classified separately using the SVM classifier. In the third scenario, selected principal
137 components are chosen from each deep feature set, and used to train SVM. Finally, in the fourth

138 scenario, all principal components selected from the four deep features sets, are fused and
139 classified using SVM.
140 Table 1. A summary of recent related studies.

2. Methods and Materials

2.1 Dataset Description

143 The details of the dataset used in this paper are available in (Zhao et al., 2020). It consists of 347
144 COVID-19 images and 397 non-COVID-19 images with several types of pathology. The length
145 of CT images ranges from 153 to 1853 pixels with a mean of 491 pixels, whereas the width ranges
146 from 124 to 383 pixels with a mean of 1485 pixels. Samples of the CT images for COVID-19 and
147 non-COVID-19 cases available in the dataset are shown in Figure 1.
148 Figure 1. Samples of CT images from the dataset; (a) – (d) COVID-19 CT images and (e) – (h)
149 non-COVID-19 CT images.

2.2 Convolutional Neural Networks

151 The convolutional neural network (CNN) is a class of DL techniques that are extensively employed
152 for solving issues regarding health informatics (Jin et al., 2020b) and in specific classification and
153 diagnosis of medical images (Ravì et al., 2016). Thus, different CNN architectures were employed
154 in this paper. A CNN consists of several layers such as convolutional layers, pooling layers, and
155 fully connected (fc) layers. In the convolutional layer, a number of filters of a specific size are
156 convolved with the portion of the input image equivalent to the same size of the filter. Afterward,
157 a feature map is generated which corresponds to the location of the features in the original image.
158 This feature represents the spatial information of the pixel values of the input image. Next, the
159 pooling layer down-samples the huge dimension of the feature map. Finally, the FC layer performs
160 the classification process like the traditional artificial neural network (ANN). The key benefit of
161 CNN over the traditional ANN, is that the feature map is now a spatial demonstration of the
162 variables existing in the dataset, and the features produced will be significant and descriptive
163 (Greenspan, Van Ginneken & Summers, 2016). Moreover, CNN does the feature extraction
164 process itself, which is not the case in traditional ANN where handcrafted feature extraction
165 techniques should be used separately then used as input to ANN (Valliani, Ranti & Oermann,
166 2019). Four CNNs corresponding to the state-of-the-art networks used for medical images are
167 employed in this paper. These networks include AlexNet, GoogleNet, ResNet-18, and ShuffleNet
168 architectures. These networks have not yet been used to detect and classify COVID-19 in the
169 literature.

2.2.1 AlexNet

171 The AlexNet CNN network was introduced in 2012 by the authors in (Krizhevsky, Sutskever &
172 Hinton, 2012). It was proposed after LeNet was created by Yann LeCun (LeCun et al., 1998). The
173 authors created the AlexNet model, which won the ImageNet Large-Scale Visual Recognition
174 Challenge in 2012. The structure of AlexNet contains 23 layers, which includes five convolutional
175 layers, five rectified linear unit (ReLu) layers, two layers for normalization, three pooling layers,
176 three fc layers, a layer of probabilities using softmax units, and lastly one layer for classification
177 terminating in 1000 neurons for 1000 categories. AlexNet CNN was initially trained with
178 ImageNet data, using more than 15 million labeled images with almost 22,000 classes.

2.2.2 GoogleNet

180 Researchers at Google designed the GoogleNet architecture in 2014. Its construction is much
181 deeper than AlexNet and consists of 22 layers. GoogleNet architecture is based on the inception

182 block, which considers dropping the number of parameters in a CNN. Therefore, it contains a
183 lower number of parameters, almost 12 times lower than AlexNet, which leads to faster
184 convergence. The GoogleNet structure won the ImageNet Large-Scale Visual Recognition
185 Challenge (ILSVRC14) in 2014, for achieving the best performance (Szegedy et al., 2015). The
186 main module of this structure is nine inception blocks that occur at every layer of the GoogleNet.
187 These blocks are scaled upon each other with a maximum pooling layer. Lastly, an fc layer exists
188 just before the output layer.

2.2.3   ShuffleNet

190 Recognizing important patterns from images, requires the construction of bigger and deeper
191 convolution neural networks. Producing accurate CNNs regularly requires a huge number of layers
192 and channels, which leads to a very high computation cost (Girshick et al., 2014). To overcome
193 this problem, an extremely efficient CNN was proposed called ShuffleNet. ShuffleNet was first
194 introduced in 2018 by Zhang et al in (Zhang et al., 2018), for applications that have very limited
195 computing power such as mobile applications. This novel network structure presents two new
196 processes namely; pointwise group convolution and channel shuffle. The former procedure uses
197 1×1 convolution to lower the computation cost, while obtaining acceptable accuracy compared to
198 the state-of-the-art CNN such as ResNet and Xception, which turns out to be less effective in
199 tremendously small networks due to the costly dense 1×1 convolutions. The later operation assists
200 the data flowing across feature channels. The channel shuffle operation lets a group of
201 convolutions attain input data from several groups, where the output/input channels are completely
202 associated. Specifically for the feature map produced in the previous group layer, the channels in
203 each group are split into separate subgroups. Afterward, each group of the preceding layer is fed
204 by a separate subgroup.

2.2.4   ResNet-18

206 One of the recent architectures that are commonly used for medical imaging applications is
207 ResNet. It received the first place in ILSVRC and the COCO 2015 competition in ImageNet
208 Detection, ImageNet localization, Coco detection, and Coco segmentation. The main building
209 block in ResNet is the residual block introduced by He et al. (He et al., 2016). Such block offers
210 short cut connections within the convolution layers, which allow the network to skip several
211 convolution layers at a time. In other words, the residual block offers two choices, it can
212 accomplish a set of functions on the input, or it can pass this phase completely. Therefore, the
213 ResNet structure is considered more efficient and produces better performance than other CNNs
214 such as AlexNet and GoogleNet. In this paper, ResNet-18 was employed which contains 17
215 convolutional layers and one fully connected layer at the end of the network.

2.3   Proposed MULTI-DEEP CAD System

217 A novel CAD system "MULTI-DEEP", for detecting COVID-19 and distinguishing it from non-
218 COVID-19 is proposed. The proposed CAD is based on the fusion of multiple CNNs. It employs
219 several CNNs that are different from those used in the literature. It consists of four approaches;
220 end-to-end classification, deep feature extraction, principal component selection, and detection. In
221 the end-to-end DL classification approach, four CNNs of different architectures were constructed
222 with the images available in the dataset. Each of the CNNs performs the classification and
223 diagnosis procedure individually. Afterward, in the deep feature extraction approach, valuable
224 features were extracted from each deep CNN separately. Next, in the principal component
225 selection approach, PCA was applied on each deep feature set individually, and the optimal
226 number of principal components was selected for each feature set separately. In the detection

227 approach, SVM classifiers were constructed to test the performance of the proposed CAD system.
228 The four approaches of the proposed CAD system were tested with four different scenarios, to
229 verify that the fusion of multiple CNNs was capable of enhancing the accuracy of the CAD system
230 and it was better than using individual CNNs. Thus, the first three scenarios were done to examine
231 the performance of the CAD using individual CNNs either by end-to-end DL classification or by
232 extracting deep features from each CNN and reducing them by PCA, then using them individually
233 to train SVM classifiers. This process was done in the first three scenarios. In the first scenario,
234 four pre-trained end-to-end networks were fine-tuned to classify COVID-19 and non-COVID-19
235 cases. In the second scenario, the deep features were extracted from each CNN and then used
236 individually to construct SVM classifiers. Additionally, in the third scenario, PCA was applied on
237 each deep feature set separately, and then a chosen number of principal components was used to
238 train SVM classifiers individually. The performance of the CAD system using individual CNNs
239 (the first three scenarios) was compared with the fusion of multiple CNNs in the fourth scenario.
240 In this last scenario, the deep features were fused to examine the influence of fusing the four deep
241 features, and determine its impact on both detection accuracy and computational efficiency.
242 Moreover, PCA was performed on the fusion feature set. Figure 2 illustrates the MULTI-DEEP
243 CAD system. Note that, a pre-processing step was done to resize all the images to be equal to the
244 required input size of the CNN architectures, as each CNN requires a different input image size.
245 Figure 2. The proposed different scenarios for the MULTI-DEEP CAD system.

2.3.1 End-to-End Deep Learning Approach
247 In this approach, the transfer learning method was employed. Transfer learning is a technique that
248 allows a pre-trained CNN that has been trained with a huge dataset containing thousands of images
249 like ImageNet, to be used in a similar classification task. In other words, a pre-trained CNN which
250 was previously learned with a similar task to the one at hand is used for a new medical task
251 (Zemouri, Zerhouni & Racoceanu, 2019), (Greenspan, Van Ginneken & Summers, 2016). Transfer
252 learning is an important step to solve convergence and overfitting issues that could happen during
253 the first few epochs. Therefore, four pre-trained CNNs of different architectures were employed.
254 These networks include AlexNet, GoogleNet, ResNet-18, and ShuffleNet architectures. A
255 summary of the architecture of the four networks is shown in Table 2. These CNNs were used
256 individually to detect COVID-19 and distinguish them from non-COVID-19 cases.
257 Table 2. A summary of the architecture of the four pre-trained CNNs used.

2.3.2 Deep Feature Extraction Approach
259 Pre-trained CNNs can either be trained from images to perform classification tasks, or be used as
260 a feature extractor where significant deep features are extracted from the FC layers of the CNNs.
261 In this approach, as an alternative of employing the CNNs as classifiers, deep features are pulled
262 out from the "fc7" of the fine-tuned AlexNet, dropout layer named "pool 5 drop 7×7 s1" of the
263 pre-trained GoogleNet, "node 200" of pre-trained ShuffleNet, and "global average pooling 2D
264 layer" (fifth pooling layer) of the fine-tuned ResNet-18. The number of deep features generated
265 from each CNN is 544, 4096, 1024, and 512 for SuffleNet, AlexNet, GoogleNet, and ResNet-18,
266 respectively.

2.3.3 Principal Component Selection Approach
268 The deep features extracted in the previous approach are of large dimension. Therefore, in this
269 approach PCA is employed to reduce the dimension of the feature space vector. PCA is a popular
270 feature reduction approach, used to compress the data size by operating a covariance investigation
271 among variables of a dataset. It drops out the sum of observed variables to lessen the principal

272 components. Such principle components demonstrate the variance of the observed variables
273 (Smith, 2002). First, PCA is employed on each deep feature individually. Afterward, the optimal
274 number of principal components for each deep feature set is chosen using a forward selection
275 procedure. Finally, these selected numbers of principal components are fused and used to train
276 SVM classifiers in the next approach.

2.3.4   Detection Approach

278 In this approach, the detection process is performed using four different scenarios. The first
279 scenario presents the detection of COVID-19 using the four pre-trained CNNs. The other three
280 scenarios include the use of a cubic SVM classifier, for detecting COVID-19 and distinguishing it
281 from non-COVID-19 cases. In the second scenario, each deep feature extracted in the deep feature
282 extraction approach is used to construct and train four cubic SVM classifiers individually. In the
283 third scenario, the principal components selected from the PCA applied to each deep feature set
284 are used to train the cubic SVM classifier individually. In the fourth scenario, the effect of fusing
285 the four deep features is investigated to determine its impact on both detection accuracy and
286 computational efficiency. Moreover, PCA is applied to the fused features to reduce their dimension
287 and examine if using PCA on the fused features will enhance the performance of the SVM
288 classifier.

3.   Experimental Set up

3.1 Parameter Setting

291 Several parameters are tuned for the pre-trained CNNs. The number of epochs is 20, and the initial
292 learning rate for the four CNNs is 10^{-4}. The mini-batch size and validation frequency are 10 and
293 4, respectively. The L-regularization is 0.0005. Further parameters are kept unchanged. These
294 configurations are to confirm that the parameters are fine-tuned for the detection of medical data.
295 Stochastic gradient descent with momentum (SGDM) is used for optimization. To validate the
296 performance of the proposed MULTI-DEEP CAD system, 5-fold cross-validation is used. Note
297 that the kernel function used for the SVM classifier is cubic, as it achieved the best performance.

3.2 Augmentation

299 Augmentation is an important process done to increase the size of the dataset, since training the
300 classification model with a small amount of data might over-fit (Ravì et al., 2016). In other words,
301 the classification of the model will memorize the details of the training set, and will not perform
302 well on testing sets. In this paper, the augmentation techniques used to create new lung CT images
303 from the training data, are flipping, translation, and scaling. Each CT image was translated in x
304 and y-directions with pixel range (-30 – 30). Moreover, each original image was flipped. They
305 were also scaled with a scaling range (0.9 – 1.1).

4.   Performance Evaluation

307 The performance of the proposed MULTI-DEEP CAD system is measured with several metrics
308 such as accuracy, sensitivity, specificity, precision, F1-score, and area under receiving operating
309 characteristics (AUC). The equations used to calculate such metrics are shown below (1-5).

$$Accuracy = \frac{TP + TN}{TN + FP + FN + TP} \tag{1}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{2}$$

312 $$Specificity = \frac{TN}{TN + FP}$$ (3)

313 $$Precision = \frac{TP}{TP + FP}$$ (4)

314 $$F1 - Score = \frac{2 \times TP}{2 \times TP + FP + FN}$$ (5)

315 Where TP is the number of COVID-19 images that are correctly classified, TN is the number of
316 non-COVID-19 images that are correctly classified. FP is the number of non-COVID-19 images
317 that are misclassified as COVID-19. FN is the number of COVID-19 images that are misclassified
318 as non-COVID-19.

5. Results

319

320 This study proposed a novel CAD system called MULTI-DEEP, based on the fusion of multiple
321 CNNs to detect and classify COVID-19 and non-COVID-19 cases. The framework was composed
322 of four different scenarios. In this section, each scenario is illustrated and the results of the four
323 scenarios are discussed.

5.1 Scenario I

324

325 In this scenario, four pre-trained CNNs were used to detect COVID-19 and non-COVID-19 cases
326 individually. The results of this scenario are shown in Table 3.
327 Table 3. Performance metrics of the four pre-trained CNNs of scenario I.
328 The highest performance was achieved by ResNet-18 with an accuracy of 78.29%, an AUC of
329 0.8386 (83.86%), a sensitivity of 0.769 (76.9%), a specificity of 0.799 (79.9%), a precision of 0.81
330 (81%), and an F1-score of 0.789 (78.9%). Whereas, the accuracies and AUC of other CNNs
331 covered the ranges (71.99% - 75.89%) and (0.7754 - 0.8166) respectively. In addition, the
332 sensitivity, specificity, precision, and F1-score fluctuated within the ranges (0.687 - 0.762), (0.724
333 - 0.768), (0.715 - 0.81), (0.728 - 0.758) respectively. Additionally, the execution time of ResNet-
334 18 was 19 m and 31 s, which proved to be the best performance compared to other networks.

5.2 Scenario II

335

336 In this scenario, deep features were extracted from each pre-trained CNN, and were used to train
337 SVM classifiers individually. The performance results of the SVM classifiers, constructed with
338 deep features extracted from AlexNet, GoogleNet, ResNet-18, and ShuffleNet, are shown in Table
339 4.
340 Table 4. Performance metrics of SVM classifiers trained individually with each deep feature
341 extracted from the pre-trained CNNs.
342 The deep features of ResNet-18 CNN achieved the highest performance compared to other CNN
343 deep features, achieving an accuracy and AUC of 92.5% and 0.97 (97%), respectively, as shown
344 in Fig. (3-a). Moreover, the other scores attained were the highest values as well. The sensitivity,
345 specificity, precision, F1-score, AUC, and the execution time reached 0.933 (93.3%), 0.918
346 (91.8%), 0.916 (91.6%), 0.925 (92.5%), and 1.947 s respectively, as demonstrated in Table 4.
347 Figure 3 shows the receiver operating characteristics (ROC) curve, and the computed AUC for the
348 SVM classifier constructed with the deep features of ResNet-18 and ShuffleNet, as these networks
349 achieved the highest results.
350 Figure 3. ROC curve for SVM classifier trained with deep features extracted from (a) pre-trained
351 ResNet-18 and (b) pre-trained ShuffleNet.

5.3 Scenario III

352

353   In Scenario III, PCA was applied to each deep feature extracted from each pre-trained CNN. Then,
354   a selected number of principal components from each deep feature set was used to train SVM
355   classifiers individually. The number of principal components was selected in a forward sequential
356   procedure. The accuracy of the constructed SVM classifiers with a number of principal
357   components is shown in Figure 4.

358   Figure 4. The performance of the SVM classifier constructed with various principal components.

359   It was noticeable from Fig. 4 that only 50 principal components for the ResNet-18 and the
360   ShuffleNet deep features provide the highest accuracy. Therefore, 50 components were selected
361   for deep features extracted from ResNet-18 and the ShuffleNet. However, for the GoogleNet deep
362   features, the highest accuracy was achieved using 50 and 100 principal components. While the
363   highest accuracy achieved for the AlexNet deep features, required 100 and 150 principal
364   components. Therefore, four-fused feature sets were generated to train the SVM classifier and then
365   compared to determine the best combination, which had an impact on the accuracy. Table 5 shows
366   a comparison between the accuracies achieved by SVM classifiers, trained with the four feature
367   sets corresponding to different combinations of selected principal components chosen from the
368   four pre-trained CNNs.

369   Table 1. Performance metrics of the SVM classifiers constructed with the four feature sets formed
370   from different combinations of the selected number of principal components.

371   Feature Set (1) contained 50 Components from each of ResNet-18, GoogleNet, and ShuffleNet,
372   and 100 components from AlexNet. Whereas, Feature set (2) had 50 components from each
373   ResNet-18, GoogleNet, and ShuffleNet, and 150 components from AlexNet. Feature Set (3)
374   composed of 50 components from each of ResNet-18 and ShuffleNet, 100 components from
375   GoogleNet, and 150 components from AlexNet. Feature Set (4) had 50 components from each of
376   ResNet-18 and ShuffleNet, and 100 components from GoogleNet and AlexNet. It was clear from
377   Table 5 that the highest SVM performance was achieved when constructed using Feature set (2).
378   Feature set (2) yields to an accuracy of 0.94 (94%), an AUC of 0.98 (98%), a sensitivity of 0.949
379   (94.9%), a specificity of 0.932 (93.2%), a precision of 0.93 (93%), and an F1-score of 0.94 (94%).

380   **5.4 Scenario IV**

381   In this scenario, the four deep features extracted from each pre-trained CNN were fused to study
382   the influence of this fusion on the performance of the SVM classifier, and compare it with the
383   fusion of the selected principal components of scenario III. The results of this fusion compared to
384   scenario III were shown in Table 6.

385   Table 2. Performance metrics of scenario IV compared to scenario III and deep features fusion
386   with PCA.

387   Scenario IV results were also compared to the fusion of the four deep features after applying PCA,
388   as shown in Table 6. The Table shows that feature set (2) of scenario III (corresponding to the
389   fusion of 50 principal components from each ResNet-18, GoogleNet, and ShuffleNet, and 150
390   principal components from AlexNet), has better performance than the fusion of all deep features,
391   after applying PCA on the fused deep features. The scores of scenario III achieved an accuracy of
392   94%, a sensitivity of 0.949 (94.9%), a specificity of 0.932 (93.2%), an AUC of 0.98 (98%), a
393   precision of 0.93 (93%), and an F1-score of 0.94 (94%). However, the scores of the fusion of deep
394   features achieved an accuracy of 92.2%, a sensitivity of 0.93 (93%), a specificity of 0.918 (91.8%),
395   an AUC of 0.97 (97%), a precision of 0.916 (91.6%), and an F1-score of 0.923 (92.3%). Scenario
396   IV achieved an accuracy of 94.7%, which was slightly higher than 94% of the accuracy achieved
397   in scenario III, as shown in Table 6. This means that scenario III (corresponding to the fusion of
398   50 principal components from each ResNet-18, GoogleNet, and ShuffleNet, and 150 principal

399 components from AlexNet) had almost the same performance of scenario IV, but with more
400 efficient computational cost as the execution time of scenario III was 2.368s, which was lower
401 than that of scenario IV (33.765 s). A comparison between the highest achieved performances in
402 each of the four scenarios is illustrated in Figure 5.
403 Figure 5. A comparison of the highest accuracy achieved for each scenario.

404 ## 6. Discussion
405 COVID-19 has led to the worst worldwide health crisis in recent history. The rapid diagnosis of
406 COVID-19 is essential to control the spread of such a disease, and the current pandemic situation.
407 The common method used to diagnose COVID-19 cases is the RT-PCR test, however, it has
408 several limitations. Among these limitations are low sensitivity, insufficient availability of such a
409 test, and the crucial need for a well-equipped laboratory for analysis (which is a key challenge
410 especially in developing countries), delayed responses and the high cost of this test. All these
411 limitations exposed the need for a quick alternative that enables the front-line specialists to achieve
412 a fast and precise diagnosis.

413 CT imaging technique has proven to be efficient in visualizing lungs, and may facilitate the
414 primary detection of suspected individuals. However, CT imaging alone was not enough for an
415 accurate diagnosis. This was due to the similarity between configurations of COVID-19 and other
416 types of pneumonia, which might confuse radiologists trying to distinguish COVID-19.
417 Nonetheless, CAD systems based on DL techniques have shown to be more capable of
418 discriminating between COVID-19 and non-COVID-19 cases, and can achieve an accurate
419 diagnosis sooner than the clinical exam, which accordingly reduces the time needed for disease
420 control (Butt et al., 2020), (Shi, Han & Zheng, 2020).

421 The ease of accessibility of CT imagining techniques makes CT-imaging a preferable tool for
422 corona virus diagnosis in many counties such as China (He et al., 2020). Moreover, the cost of the
423 diagnoses is an important factor for choosing the tool for the novel virus diagnosis. The relative
424 low cost of CT in China makes it as well a favorable diagnostic tool (He et al., 2020). RT-PCR
425 test is considered an expensive diagnostic tool in many countries including developing countries
426 in Africa such as South Africa, Algeria, Egypt, Morocco and Tunisia. This is due to lack of
427 availability of such test which increase its price (Kobia & Gitaka, 2020). The presently available
428 RT-PCR kits are variable, there are more than 7 platform used for the novel corona virus diagnosis
429 (Al-Tawfiq & Memish, 2020), (van Kasteren et al., 2020). Such tests may result in initial false
430 negative outcomes (Long et al., 2020). Examples of such kits are Altona Diagnostics, BGI, CerTest
431 Biotec, KH Medical, Primer Design, R-Biopharm AG, and Seegene (van Kasteren et al., 2020).
432 RT-PCR platforms are presenting sensitivities ranging between 45% and 60%; therefore,
433 specifically in the initial course of an infection, repeat testing may be essential to achieve a
434 diagnosis. This is difficult to attain with the global shortage of testing kits specially in developing
435 counties (Al-Tawfiq & Memish, 2020). Moreover, RT-PCR tests may take few hours to complete,
436 however they are restricted by the time needed to transfer and arrange sample for testing.
437 Nowadays, due to the increase in the demand of such test, the laboratories have been flooded with
438 samples leading to significant delays in diagnostic duration (Kim, Hong & Yoon, 2020), (Ai et al.,
439 2020), (Young et al., 2020). In addition, the availabilities of such labs are limited in some
440 developing countries like Egypt, in contrast to the CT machines, which are located in many

441 hospitals and scan centers. The authors in (Shen et al., 2020) showed that CT results are obtainable
442 in time which are lower than PCR test. Also, (Ai et al., 2020) compared the speed and accuracy
443 sensitivity of TaqMan One-Step RT-PCR Kits with CT and the results showed that CT is more
444 sensitive and faster than PCR kit used. This make the use CT more preferable specially when used
445 with artificial intelligence techniques such DL (Kobia & Gitaka, 2020). Several studies showed
446 that using CAD systems based on DL and CT images are considered a rapid, more accurate and
447 sensitive than the PCR test. This is because when a CT image enters the CAD, it is automatically
448 classified to COVID and non COVID in less than an hour with better accuracy and sensitivity than
449 the PCR test (Das et al., 2020), (Manigandan et al., 2020, p. 19), (Song et al., 2020), (Maghdid et
450 al., 2020).

451 In this paper, we proposed a fast solution, that may serve as an alternative to the RT-PCR test. The
452 proposed solution introduced a novel CAD system, based on the fusion of multiple CNNs trained
453 with CT images. This study is a crucial trial comprising a simple to construct, low cost, and
454 automatic CAD system that can achieve a high accuracy, based on the fusion of multiple CNNs.
455 During these difficult days of the worldwide pandemic, the proposed MULTI-DEEP CAD has
456 great potential to be used as a tool for COVID-19 detection. It was verified that CT provides a
457 convenient and effective method to primary identify suspicious cases (Xu et al., 2020), (Long et
458 al., 2020). Thus, the proposed system can detect COVID-19 cases early, thereby avoiding the fast
459 spread of the disease. The proposed CAD is a CNN-based model. CNN methods provided that the
460 characteristic lung lesions exist at the time of testing are more capable of distinguishing between
461 COVID and non-COVID-19 cases than a manual diagnosis from CT images (Butt et al., 2020).
462
463 The outperformance of CNN was proven in the related studies by various authors (Bai et al., 2020),
464 (Chen et al., 2020), (Ardakani et al., 2020), who compared the performance of their CAD system
465 based on deep learning techniques, with that of a trained radiologist without the help of a CAD.
466 These studies indicated that the performance of DL based CAD systems outperformed manual
467 diagnosis by a trained radiologist without the help of a CAD. The authors in (Bai et al., 2020)
468 showed that their DL based CAD system achieved higher test accuracy (96% vs. 85%), sensitivity
469 (95% vs. 79%), and specificity (96% vs. 88%) than radiologists. Whereas in (Ardakani et al., 2020)
470 the performance of the radiologist was lower than the authors' proposed DL-based CAD system,
471 with a sensitivity of (89.21% vs. 98.04%), specificity of (83.33% vs. 98.04%), and accuracy of
472 (86.27% vs. 99.5%). On the other hand, in (Chen et al., 2020), the authors showed that their DL
473 based CAD system can reduce the manual radiologist's diagnosis time by 65%.

474 As it was obvious from Fig. 5 that scenario III, corresponding to the fusion of the selected number
475 of principal components from each deep feature set, has increased the accuracy to 94% compared
476 to 92.5% achieved in scenario II using SVM classifier trained using only deep features extracted
477 from ResNet-18. Although, the accuracy was slightly improved to 94.7% when fusing all deep
478 features extracted from the four pre-trained CNNs without PCA, the execution time in scenario III
479 is 2.105 s, which was lower than the 33.765 s of scenario IV, and much lower than the 19 m 31 s
480 of scenario I. This means that the fusion of the selected principal components led to a more efficient
481 CAD system.

482 MULTI-DEEP has achieved accuracy, AUC, sensitivity, and specificity of 94.7%, 98%, 95.6%,
483 and 93.7% for identifying COVID-19 and non-COVID-19 cases. By comparing the results of the
484 proposed MULTI-DEEP CAD system with the related work presented in Table 1, it was noticed
485 that the CNN architectures used were different from those used in MULTI-DEEP. Regarding the
486 ResNet CNN architectures, it was obvious that the researchers used it with different layer
487 structures than in (Butt et al., 2020), (Li et al., 2020a), (Jin et al., 2020a), (Jin et al., 2020b), and
488 (Song et al., 2020). Butt et al. (Butt et al., 2020) used the ResNet-18 and ResNet-23 CNNs to
489 classify COVID-19 and non-COVID-19 samples, achieving an accuracy of 86.7%, which is lower
490 than that of MULTI-DEEP. On the other hand, Jin et al. (Jin et al., 2020a) performed the
491 classification using ResNet-152 CNN reaching an accuracy of 94.8%, which is almost the same
492 accuracy of MULTI-DEEP, which was constructed with a fewer number of images. It is expected
493 that the performance of the proposed MULTIDEEP will be enhanced when the number of images
494 is increased. Shi et al. (Li et al., 2020a), Song et al. (Song et al., 2020), and Jin et al. (Jin et al.,
495 2020b) used the ResNet-50 achieving an accuracy of 89.5%, 86%, and 94.8% respectively. The
496 accuracies achieved by (Li et al., 2020a) and (Song et al., 2020), are much lower than the proposed
497 MULTI-DEEP, but the accuracy attained by (Jin et al., 2020b) is almost the same as the proposed
498 method and this is because of the large amount of data employed by Jin et al. (Jin et al., 2020b).
499 Furthermore, when Amyar et al. (Amyar, Modzelewski & Ruan, 2020) and Chen et al. (Chen et
500 al., 2020) employed the U-Net for the segmenting and/or classification processes, the accuracy
501 reached 86%, and 95.2% respectively. The high accuracy achieved by Chen et al. (Chen et al.,
502 2020) is due to the huge number of images used to train their U-NETs compared to the proposed
503 MULTI-DEEP and this is one of the limitations of MULTI-DEEP. However, MULTI-DEEP
504 achieved acceptable performance, even with a smaller number of images. Zheng et al in (Zheng
505 et al., 2020) employed a U-Net for segmenting CT images, and then the authors constructed a CNN
506 with eight layers. The accuracy attained was 90.9%, which was much lower than achieved in
507 MULTI-DEEP.

508 It is quite obvious from the previous comparison that the proposed MULTI-DEEP performs
509 competitively, compared to the most recent CAD systems shown in the literature. The results of
510 MULTI-DEEP verify that the proposed MULTI-DEEP CAD system can be used to overcome the
511 limitation of using CT imaging alone to diagnose the COVID-19 virus. The satisfactory accuracy
512 achieved by the proposed model, even with its primary outcomes, unlocks the door for the
513 production of a comprehensive product by IT (information technology) solution companies. This
514 product can work mobile and appeal to the end-user, which are the radiologists or medical experts,
515 to facilitate the diagnosis process of the COVID-19. It can be made as a mobile application or
516 webpage, where a CT image is imported to the application or webpage and a decision is exported.

517 The possible scenario where this algorithm can be used, is when a patient is suffering from any
518 coronavirus symptoms. The medical expert can take a CT scan of them instead of the RT-PCR
519 test, then import the CT image into the proposed CAD system for a quicker, more accurate,
520 simpler, and less expensive decision. This decision made by the proposed CAD will assist the
521 specialist in improving the quality of patient management, even in extraordinary workload
522 situations. As mentioned before, the proposed CAD is preferable to the RT-PCR test, due to the

523 various limitations of the latter test. The proposed CAD can be tested on any suspected patient to
524 determine if they are infected with COVID-19.

525 A limitation of this study that could be considered in future research, is the relatively small number
526 of COVID-19 images used. The performance of the proposed CAD system is likely to improve
527 when the training data is increased. Besides, at the moment the proposed model can only
528 distinguish COVID-19 from non COVID-19 cases. It is essential to distinguish COVID-19 cases
529 from other types of pneumonia as well. Future research will be done to expand the model so that it
530 can correctly diagnose other types of pneumonia. One more limitation is that the performance of the
531 CAD was not compared with a trained radiologist. Future work will also focus on using multimodal
532 imaging techniques, using more types of CNNs, and finally employing a segmentation technique
533 to distinguish lung from other tissue, and comparing its performance to that of the proposed CAD.

534 7. Conclusions

535 Several studies have verified that deep learning can help radiologists to accurately diagnose
536 COVID-19. The key goal of this paper was to construct an efficient CAD system capable of
537 accurately detecting COVID-19, and differentiating it from non-COVID-19 cases. The proposed
538 MULTI-DEEP CAD system was based on the fusion of multiple CNNs. It went through four
539 experiments to achieve a more efficient CAD system. The experiments showed that fusing deep
540 features extracted from several pre-trained CNNs, improved the performance of the proposed CAD
541 system. The experiments also proved that fusing the optimal number of principal components
542 selected from each deep feature, has decreased the computational cost of the CAD to almost 32 %,
543 which lead to a more efficient CAD system. The results verified that the proposed MULTI-DEEP
544 CAD could successfully detect COVID-19 with high accuracy. Additionally, it showed its
545 competitive performance compared to similar studies. Therefore, MULTI-DEEP can be employed
546 by radiologists to facilitate the diagnosis process of COVID-19. MULTI-DEEP may assist in
547 managing the current pandemic, speed up the detection of the disease, avoid its fast spread, and
548 help doctors to improve the quality of patient management, even in extraordinary workload
549 situations. MULTI-DEEP should undergo further field testing before it can be employed by
550 radiologists. In addition, it will likely have to undergo regulatory approval by health authorities
551 before its implementation into hospitals.

552 **Funding:** This study did not receive any external funding.
553 **Conflict of interest:** The authors declare that they have no conflict of interest.
554 **References**

555 Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, Tao Q, Sun Z, Xia L. 2020. Correlation of chest
556   CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of
557   1014 cases. *Radiology*:200642.
558 Al-Tawfiq JA, Memish ZA. 2020. Diagnosis of SARS-CoV-2 infection based on CT scan vs RT-
559   PCR: reflecting on experience from MERS-CoV. *Journal of Hospital Infection* 105:154–
560   155.
561 Amyar A, Modzelewski R, Ruan S. 2020. Multi-task Deep Learning Based CT Imaging Analysis
562   For COVID-19: Classification and Segmentation. *medRxiv*.

563   Arabi YM, Murthy S, Webb S. 2020. COVID-19: a novel coronavirus and a novel challenge for
564   critical care. *Intensive care medicine*:1–4.

565   Ardakani AA, Kanafi AR, Acharya UR, Khadem N, Mohammadi A. 2020. Application of deep
566   learning technique to manage COVID-19 in routine clinical practice using CT images:
567   Results of 10 convolutional neural networks. *Computers in Biology and Medicine*:103795.

568   Attallah O, Sharkas MA, Gadelkarim H. 2020. Deep Learning Techniques for Automatic
569   Detection of Embryonic Neurodevelopmental Disorders. *Diagnostics* 10:27–49.

570   Bai HX, Wang R, Xiong Z, Hsieh B, Chang K, Halsey K, Tran TML, Choi JW, Wang D-C, Shi
571   L-B. 2020. AI Augmentation of Radiologist Performance in Distinguishing COVID-19
572   from Pneumonia of Other Etiology on Chest CT. *Radiology*:201491.

573   Butt C, Gill J, Chun D, Babu BA. 2020. Deep learning system to screen coronavirus disease 2019
574   pneumonia. *Applied Intelligence*:1.

575   Chen J, Wu L, Zhang J, Zhang L, Gong D, Zhao Y, Hu S, Wang Y, Hu X, Zheng B. 2020. Deep
576   learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution
577   computed tomography: a prospective study. *medRxiv*.

578   Das NN, Kumar N, Kaur M, Kumar V, Singh D. 2020. Automated deep transfer learning-based
579   approach for detection of COVID-19 infection in chest X-rays. *IRBM*.

580   Dong E, Du H, Gardner L. 2020. An interactive web-based dashboard to track COVID-19 in real
581   time. *The Lancet infectious diseases*.

582   Fang Y, Zhang H, Xie J, Lin M, Ying L, Pang P, Ji W. 2020. Sensitivity of chest CT for COVID-
583   19: comparison to RT-PCR. *Radiology*:200432.

584   Fujita H. 2020. AI-based computer-aided diagnosis (AI-CAD): the latest review to read first.
585   *Radiological Physics and Technology*:1–14.

586   Girshick R, Donahue J, Darrell T, Malik J. 2014. Rich feature hierarchies for accurate object
587   detection and semantic segmentation. In: *Proceedings of the IEEE conference on computer
588   vision and pattern recognition*. 580–587.

589   Grasselli G, Pesenti A, Cecconi M. 2020. Critical care utilization for the COVID-19 outbreak in
590   Lombardy, Italy: early experience and forecast during an emergency response. *Jama*.

591   Greenspan H, Van Ginneken B, Summers RM. 2016. Guest editorial deep learning in medical
592   imaging: Overview and future promise of an exciting new technique. *IEEE Transactions
593   on Medical Imaging* 35:1153–1159.

594   He J-L, Luo L, Luo Z-D, Lyu J-X, Ng M-Y, Shen X-P, Wen Z. 2020. Diagnostic performance
595   between CT and initial real-time RT-PCR for clinically suspected 2019 coronavirus disease
596   (COVID-19) patients outside Wuhan, China. *Respiratory Medicine*:105980.

597   He K, Zhang X, Ren S, Sun J. 2016. Deep Residual Learning for Image Recognition. In: *IEEE
598   Conference on Computer Vision and Pattern Recognition*. DOI: 10.1109/CVPR.2016.90.

599   Iwasawa T, Sato M, Yamaya T, Sato Y, Uchida Y, Kitamura H, Hagiwara E, Komatsu S,
600   Utsunomiya D, Ogura T. 2020. Ultra-high-resolution computed tomography can
601   demonstrate alveolar collapse in novel coronavirus (COVID-19) pneumonia. *Japanese
602   Journal of Radiology*:1.

Jin C, Chen W, Cao Y, Xu Z, Zhang X, Deng L, Zheng C, Zhou J, Shi H, Feng J. 2020a. Development and Evaluation of an AI System for COVID-19 Diagnosis. *medRxiv*.

Jin S, Wang B, Xu H, Luo C, Wei L, Zhao W, Hou X, Ma W, Xu Z, Zheng Z. 2020b. AI-assisted CT imaging analysis for COVID-19 screening: Building and deploying a medical AI system in four weeks. *medRxiv*.

van Kasteren PB, van der Veer B, van den Brink S, Wijsman L, de Jonge J, van den Brandt A, Molenkamp R, Reusken CB, Meijer A. 2020. Comparison of commercial RT-PCR diagnostic kits for COVID-19. *Journal of Clinical Virology*:104412.

Kim H, Hong H, Yoon SH. 2020. Diagnostic performance of CT and reverse transcriptase-polymerase chain reaction for coronavirus disease 2019: a meta-analysis. *Radiology*:201343.

Kobia F, Gitaka J. 2020. COVID-19: Are Africa's diagnostic challenges blunting response effectiveness? *AAS Open Research* 3.

Krizhevsky A, Sutskever I, Hinton GE. 2012. Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*. 1097–1105.

LeCun Y, Bottou L, Bengio Y, Haffner P. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86:2278–2324.

Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, Bai J, Lu Y, Fang Z, Song Q. 2020a. Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest ct. *Radiology*:200905.

Li D, Wang D, Dong J, Wang N, Huang H, Xu H, Xia C. 2020b. False-negative results of real-time reverse-transcriptase polymerase chain reaction for severe acute respiratory syndrome coronavirus 2: role of deep-learning-based CT diagnosis and insights from two cases. *Korean journal of radiology* 21:505–508.

Long C, Xu H, Shen Q, Zhang X, Fan B, Wang C, Zeng B, Li Z, Li X, Li H. 2020. Diagnosis of the Coronavirus disease (COVID-19): rRT-PCR or CT? *European journal of radiology*:108961.

Maghdid HS, Ghafoor KZ, Sadiq AS, Curran K, Rabie K. 2020. A novel ai-enabled framework to diagnose coronavirus covid 19 using smartphone embedded sensors: Design study. *arXiv preprint arXiv:2003.07434*.

Manigandan S, Wu M-T, Ponnusamy VK, Raghavendra VB, Pugazhendhi A, Brindhadevi K. 2020. A systematic review on recent trends in transmission, diagnosis, prevention and imaging features of COVID-19. *Process Biochemistry*.

Ragab DA, Sharkas M, Attallah O. 2019. Breast cancer diagnosis using an efficient CAD system based on multiple classifiers. *Diagnostics* 9:165-191.

Ragab DA, Sharkas M, Marshall S, Ren J. 2019. Breast cancer detection using deep convolutional neural networks and support vector machines. *PeerJ* 7:e6201.

Ravì D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B, Yang G-Z. 2016. Deep learning for health informatics. *IEEE journal of biomedical and health informatics* 21:4–21.

643 Shen C, Mark R, Kagetsu NJ, Becker AS, Bar-Yam Y. 2020. Combining PCR and CT testing for
644    COVID. *arXiv preprint arXiv:2006.02140*.
645 Shi H, Han X, Zheng C. 2020. Evolution of CT manifestations in a patient recovered from 2019
646    novel coronavirus (2019-nCoV) pneumonia in Wuhan, China. *Radiology* 295:20–20.
647 Smith LI. 2002. A tutorial on Principal Components Analysis Introduction.
648 Song Y, Zheng S, Li L, Zhang X, Zhang X, Huang Z, Chen J, Zhao H, Jie Y, Wang R. 2020. Deep
649    learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images.
650    *medRxiv*.
651 Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich
652    A. 2015. Going deeper with convolutions. In: *Proceedings of the IEEE conference on*
653    *computer vision and pattern recognition*. 1–9.
654 Valliani AA-A, Ranti D, Oermann EK. 2019. Deep Learning and Neurology: A Systematic
655    Review. *Neurology and therapy*:1–15.
656 Wang S, Kang B, Ma J, Zeng X, Xiao M, Guo J, Cai M, Yang J, Li Y, Meng X. 2020. A deep
657    learning algorithm using CT images to screen for Corona Virus Disease (COVID-19).
658    *MedRxiv*.
659 Xie X, Zhong Z, Zhao W, Zheng C, Wang F, Liu J. 2020. Chest CT for typical 2019-nCoV
660    pneumonia: relationship to negative RT-PCR testing. *Radiology*:200343.
661 Xu B, Xing Y, Peng J, Zheng Z, Tang W, Sun Y, Xu C, Peng F. 2020. Chest CT for detecting
662    COVID-19: a systematic review and meta-analysis of diagnostic accuracy. *European*
663    *Radiology*:1.
664 Young D, Liana T, Mujtaba G, Priscilla C, Samer I, Gunjan J, Haaris N, Phillip B, Krishna A,
665    Howard S, Kashif H, Akella C. 2020. Chest CT versus RT-PCR for Diagnostic Accuracy
666    of COVID-19 Detection: A Meta-Analysis. *Journal of Vascular Medicine & Surgery* 8:1–
667    4.
668 Zemouri R, Zerhouni N, Racoceanu D. 2019. Deep Learning in the Biomedical Applications:
669    Recent and Future Status. *Applied Sciences* 9:1526.
670 Zhang X, Zhou X, Lin M, Sun J. 2018. Shufflenet: An extremely efficient convolutional neural
671    network for mobile devices. In: *Proceedings of the IEEE conference on computer vision*
672    *and pattern recognition*. 6848–6856.
673 Zhao J, Zhang Y, He X, Xie P. 2020. COVID-CT-Dataset: a CT scan dataset about COVID-19.
674    *arXiv preprint arXiv:2003.13865*.
675 Zheng C, Deng X, Fu Q, Zhou Q, Feng J, Ma H, Liu W, Wang X. 2020. Deep learning-based
676    detection for COVID-19 from chest CT using weak label. *medRxiv*.

677

Table 1

A summary of recent related studies.

Paper	Dataset	Method	Results
Butt et al. (Butt et al., 2020)	219 COVID-19 339 Others	ResNet-18 and ResNet-23	Accuracy = 95.2% AUC = 0.996 (99.6%) Sensitivity = 98.2% Specificity = 92.2%
Li et al. (Li et al., 2020a)	468 COVID-19 2996 Others	ResNet-50	Accuracy = 89.5% AUC = 0.95 (95%) Sensitivity = 87% Specificity = 92%
Bai et al. (Bai et al., 2020)	521 COVID-19 665 others	EfficientNet-B4	Accuracy = 87% AUC = 0.9 (90%) Sensitivity = 89% Specificity = 86%
Amyar et al. (Amyar, Modzelewski & Ruan, 2020)	449 COVID-19 595 others	Two U-Nets	Accuracy = 86% AUC = 0.93 (93%) Sensitivity = 94% Specificity = 79%
Chen et al. (Chen et al., 2020)	4382 COVID-19 9369 others	U-Net++	Accuracy = 95.2% Sensitivity = 100% Specificity = 93.6%
Zheng et al. (Zheng et al., 2020)	313 COVID-19 229 Others	U-Net and CNN	Accuracy = 90.9% AUC = 0.959 (95.9%) Sensitivity = 90.7% Specificity = 91.1%
Jin et al. (Jin et al., 2020a)	496 COVID-19 1385 Others	Deeplab-v1 and ResNet-152	Accuracy = 94.8% AUC = 0.979 (97.9%) Sensitivity = 94.1% Specificity = 95.5%
Jin et al. (Jin et al., 2020b)	723 COVID-19 413 Others	U-Net and ResNet-50	Accuracy = 94.8% Sensitivity = 97.4% Specificity = 92.2%
Song et al. (Song et al., 2020)	219 COVID-19 399 Others	ResNet-50	Accuracy = 86% AUC = 0.95 (95%) Sensitivity = 96% Precision = 79%
Wang et al. (Wang et al., 2020)	325 COVID-19 740 Others	Inception and Adaboosted decision tree	Accuracy = 82.9% AUC = 0.9 (90%) Sensitivity = 81% Specificity = 84%

Table 2

A summary of the architecture of the four pre-trained CNNs used.

CNN Architecture	Number of Layers	Input Size	Output Size
AlexNet	8	227 × 227	4096 × 2
GoogleNet	22	224 × 224	1024 × 2
ResNet-18	18	224 × 224	512 × 2
ShuffleNet	50	224 × 224	544 × 2

Table 3

Performance metrics of the four pre-trained CNNs of scenario I.

	Accuracy	AUC	Sensitivity	Specificity	Precision	F1 Score	Time
AlexNet	73.21%	0.7754	0.74	0.724	0.715	0.728	19 m 43 s
GoogleNet	75.89%	0.8166	0.762	0.756	0.753	0.758	19 m 31 s
ResNet-18	**78.29%**	**0.8382**	**0.769**	**0.799**	**0.81**	**0.789**	**19 m 31 s**
ShuffleNet	71.99%	0.7903	0.687	0.768	0.81	0.744	25 m 26 s

Table 4

Performance metrics of SVM classifier trained individually with each deep feature extracted from the pre-trained CNNs.

	Accuracy (std)	AUC (std)	Sensitivity (std)	Specificity (std)	Precision (std)	F1 Score (std)	Time (sec) (std)
AlexNet	90.9% (0.002)	0.95 (0)	0.922 (0.005)	0.896 (0.003)	0.891 (0.005)	0.907 (0.003)	20.991 (3.066)
GoogleNet	89.2% (0.004)	0.95 (0)	0.914 (0.029)	0.86 (0.009)	0.849 (0.006)	0.881 (0.016)	3.867 (0.274)
ResNet-18	**92.5%** (0.005)	**0.97** (0)	**0.933** (0.005)	**0.918** (0.07)	**0.916** (0.007)	**0.925** (0.006)	**1.947** (0.25)
ShuffleNet	91.1% (0.002)	0.98 (0.001)	0.919 (0.003)	0.904 (0.004)	0.902 (0.005)	0.911 (0.003)	2.54 (0.168)

Table 5

Performance metrics of the SVM classifiers constructed with the four feature sets formed from different combinations of the selected number of principal components.

	Accuracy (std)	AUC (std)	Sensitivity (std)	Specificity (std)	Precision (std)	F1 Score (std)	Time (sec) (std)
Feature Set (1) (250 Features)	93.4% (0.002)	0.98 (0.001)	0.945 (0.006)	0.922 (0.003)	0.919 (0.004)	0.933 (0.003)	2.071 (0.269)
Feature Set (2) (300 Features)	**94%** (0.002)	**0.98** (0.001)	**0.949** (0.001)	**0.932** (0.001)	**0.93** (0.001)	**0.94** (0.001)	**2.105** (0.001)
Feature Set (3) (350 Features)	92.6% (0.003)	0.972 (0.005)	0.935 (0.006)	0.916 (0.005)	0.913 (0.005)	0.924 (0.004)	2.002 (0.194)
Feature Set (4) (300 Features)	93% (0.003)	0.98 (0.001)	0.935 (0.005)	0.926 (0.005)	0.925 (0.006)	0.93 (0.001)	2.088 (0.293)

Table 6

Performance metrics of scenario IV compared to scenario III and deep features fusion with PCA.

	Accuracy (std)	AUC (std)	Sensitivity (std)	Specificity (std)	Precision (std)	F1 Score (std)	Time (sec) (std)
Fusion with PCA	92.2% (0.001)	0.97 (0)	0.93 (0.001)	0.918 (0.005)	0.916 (0.006)	0.923 (0.003)	2.368 (0.354)
Feature Set (2) (Scenario III)	94% (0.002)	0.98 (0.001)	0.949 (0.001)	0.932 (0.001)	0.93 (0.001)	0.94 (0.001)	2.105 (0.001)
Fusion Without PCA (Scenario IV)	**94.7%** (0.004)	**0.98** (0.001)	**0.956** (0.007)	**0.937** (0.005)	**0.934** (0.006)	**0.945** (0.003)	**33.765** (1.46)

Figure 1

Figure 1

Samples of CT images from the dataset; (a) - (d) COVID-19 CT images and (e) - (h) non-COVID-19 CT images.
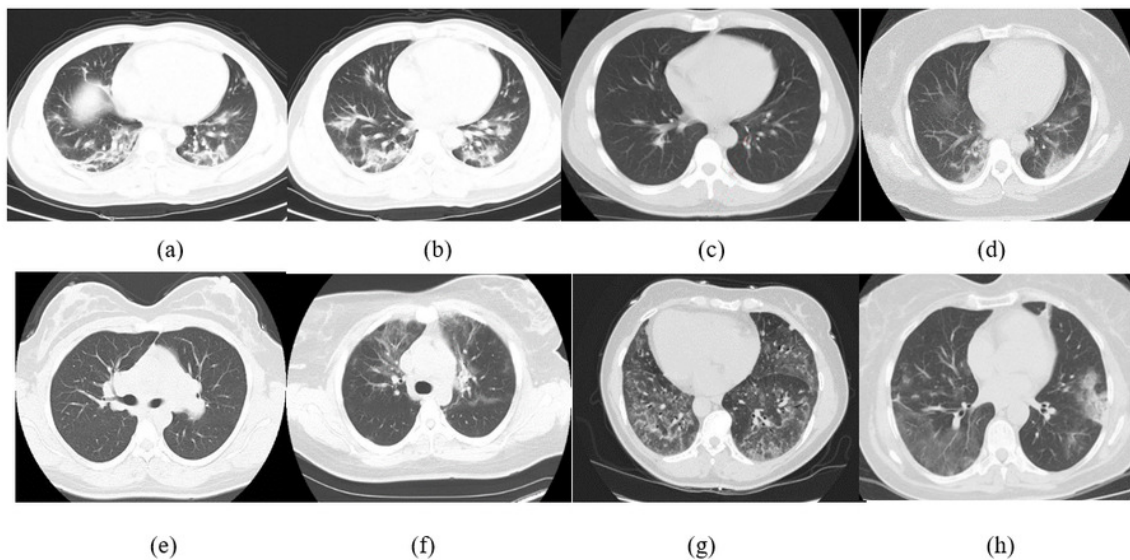
Figure 2

Figure 2

The proposed different scenarios for the MULTI-DEEP CAD system.
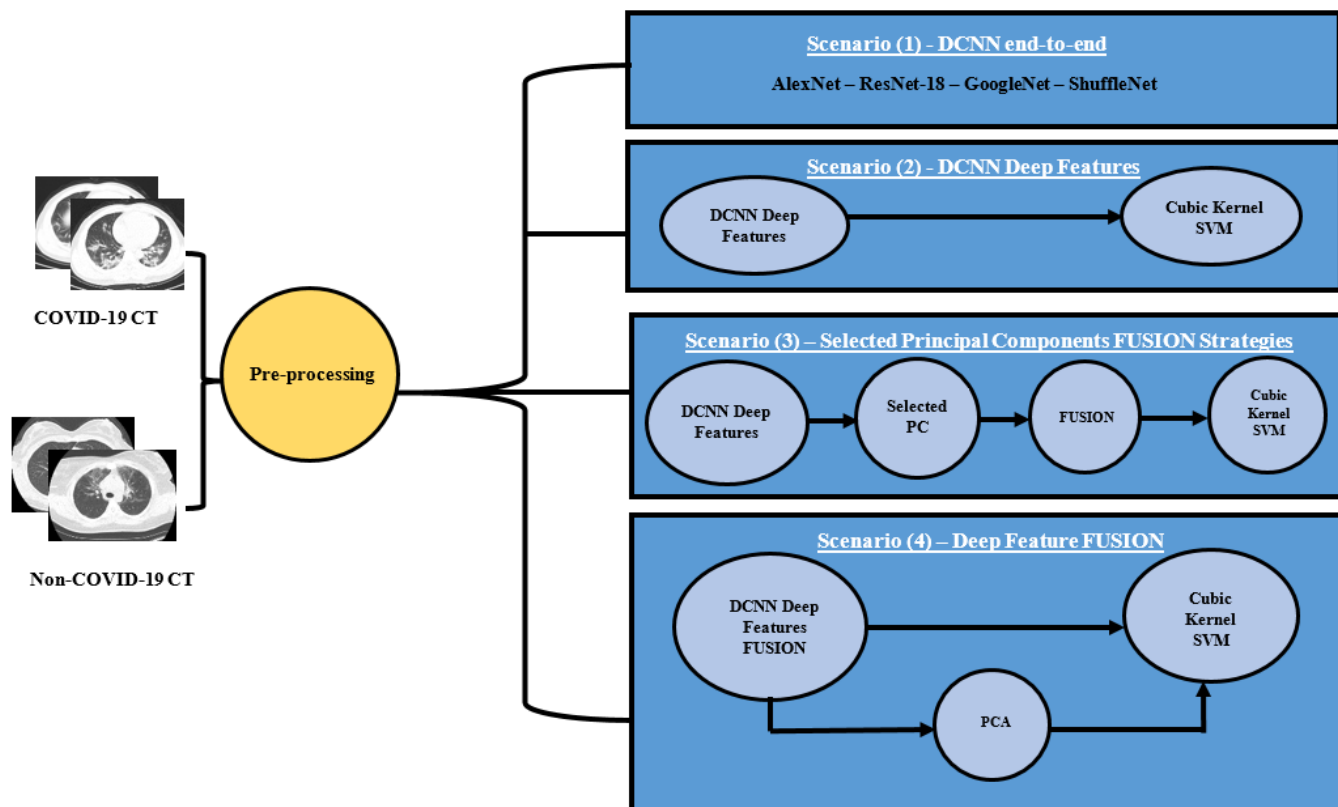
Figure 3

Figure 3

ROC curve for SVM classifier trained with deep features extracted from (a) pre-trained ResNet-18 and (b) pre-trained ShuffleNet.

(a)   (b)
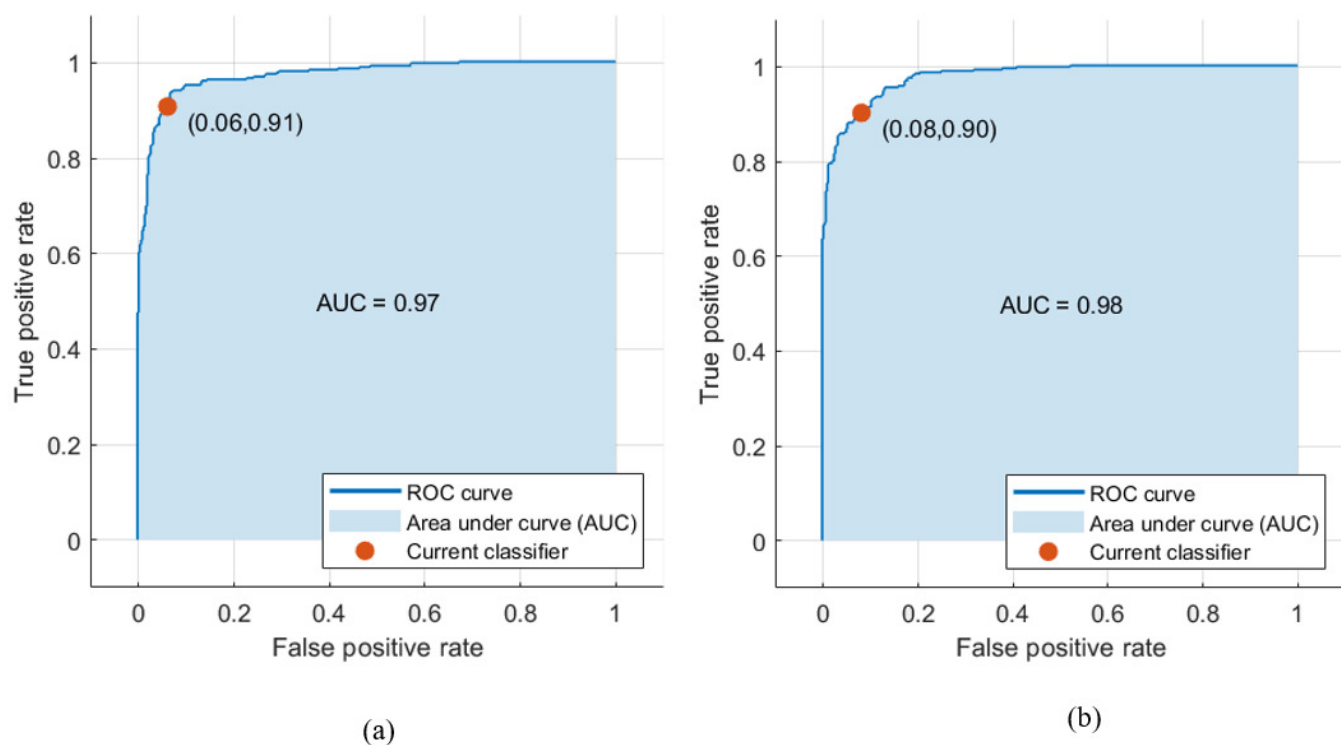
Figure 4

Figure 4

The performance of the SVM classifier constructed with various principal components.
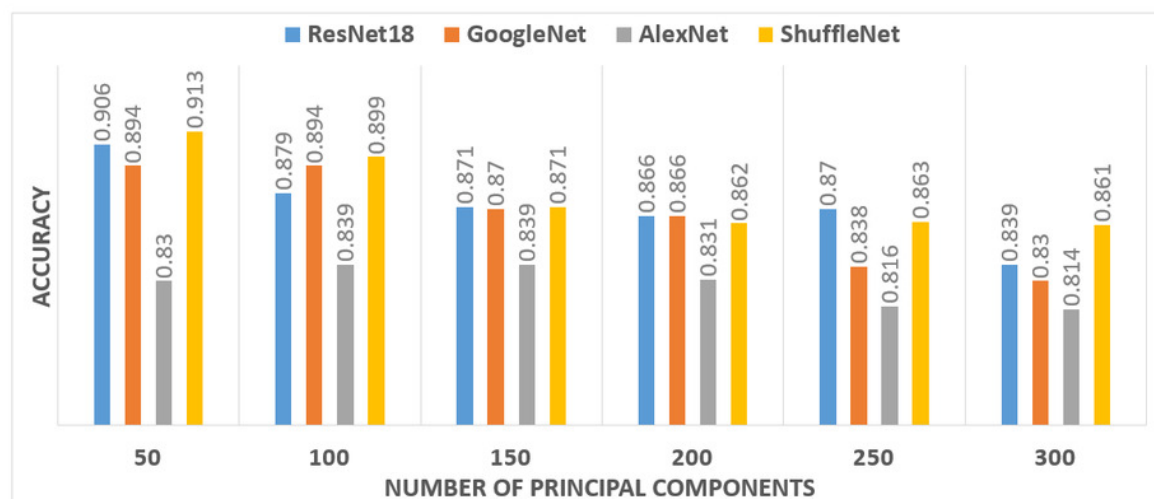
Figure 5

Figure 5

A comparison of the highest accuracy achieved for each scenario.