

Diagnosing COVID-19 in Chest X-ray Images based on Deep Learning: Transfer Learning versus Deep Features Extraction

Mohammad I. Daoud¹, Wajdi Elmuhktadi¹, Mohammad Faidi¹, Yara Alrahahleh¹, Samir Abdel-Rahman¹, Aamer Al-Ali¹, Baha A. Alsaify², and Rami Alazrai¹

¹*Department of Computer Engineering, German Jordanian University, Amman, Jordan*

²*Department of Network Engineering and Security, Jordan University of Science and Technology, Irbid, Jordan*

{mohammad.aldaoud, w.elmuhtadi, m.faidi, y.alrahahleh, s.abdelrahman, a.alali1, rami.azrai}@gju.edu.jo
and baalsaify@just.edu.jo

Abstract—Chest X-ray (CXR) images provide an effective modality for detecting COVID-19 infections. Nevertheless, the interpretation of CXR images is challenging and operator-dependent task. Several studies proposed the use of pretrained convolutional neural network (CNN) models to classify CXR images with the goal of detecting COVID-19 infections. In fact, the classification of CXR images using the pretrained CNN models is essentially performed using two approaches, namely the transfer learning approach and deep features extraction approach. This study aims to compare the performance of these two approaches to classify CXR images as COVID-19, pneumonia, and normal. Three pretrained CNN models, namely the AlexNet, VGG19, and ResNet50 CNN models, have been utilized. Furthermore, a balanced dataset of CXR images is used to perform the analysis, where this dataset includes 1,228 COVID-19 CXR images, 1,228 pneumonia CXR images, and 1,228 normal CXR images. For the three pretrained CNN models, the deep features extraction approach achieved better classification results compared with the transfer learning approach. Moreover, the results show that the ResNet50 CNN model obtained the highest classification performance based on the transfer learning approach and the deep features extraction approach. The highest macro-averaged sensitivity, specificity, and F1 score values, which have been achieved using the deep features extraction approach and the ResNet50 CNN model, are equal to 93.7%, 96.9%, and 93.7%, respectively.

Index Terms—COVID-19 detection, convolutional neural networks, chest X-ray image classification, transfer learning, deep features extraction

I. INTRODUCTION

The fast spread of the Coronavirus Disease 2019 (COVID-19) has led to extensive global-scale consequences that affected different sectors, including the medical sector, economical sector, and educational sector. Up to February 2022, more than 400 million persons around the globe were diagnosed with COVID-19 and around 6 million have died due to the disease [1]. Detecting COVID-19 infections at early stages is crucial to minimize the spread of the disease and perform medical treatment, if needed, to reduce the severity of the disease.

Currently, the reverse transcription polymerase chain reaction (RT-PCR) test is the main detection tool for detecting

COVID-19 infections [2]. However, the RT-PCR test has some potential limitations, including the long time required to obtain the results and the low sensitivity in some mild infections [3]. Chest X-ray (CXR) images, which can be employed to detect the lung variations related to the disease, provide an important detection tool to enhance the diagnosis of COVID-19 [4], [5]. Nevertheless, the manual reading and interpretation of CXR images is an operator-dependent task that requires experienced radiologists.

Many researchers proposed the use of deep learning technology, and particularly convolutional neural network (CNN) models, to develop automated systems for analyzing and classifying CXR images. The survey study by Alghamdi et al. [6] provides detailed review about the methods proposed to apply deep learning technology for classifying CXR images with the goal of detecting COVID-19 infections. In general, there are two important classification approaches, namely the transfer learning approach and deep features extraction approach, that enable the use of pretrained CNN models to classify CXR images. The transfer learning approach aims to fine-tune and adapt the pretrained CNN models to carry out the task of classifying CXR images. Examples of the studies that used the transfer learning approach for classifying CXR images include the works by Nayak et al. [7], Rahaman et al. [8], and Rehman et al. [9]. Recently, Daoud et al. [10] proposed a hierarchical classification strategy to improve the capability of the transfer learning approach to classify CXR images based on pretrained CNN models.

The deep features extraction approach aims to employ the pretrained CNN models to extract deep features for characterizing the CXR images, and then analyze the extracted features using a computer classifier. The deep features extraction approach has been used in some previous studies. For example, Turkoglu [11] employed the pretrained AlexNet CNN model to extract deep features from the CXR images and utilized a Support Vectors Machine (SVM) [12] classifier to classify the extracted features. Moreover, Sitaula and Aryal [13] proposed the use of the Bag of Visual Words method to improve the capability of the deep features extraction approach to classify

CXR images.

The current study aims to compare the capability of the transfer learning approach and the deep features extraction approach to classify CXR images as COVID-19, pneumonia, and normal. Three pretrained CNN models have been considered, where these models are the AlexNet [14], VGG19 [15], and ResNet50 [16] models. Furthermore, a balanced dataset of CXR images has been used to perform the analysis, where this dataset includes 1,228 COVID-19 CXR images, 1,228 pneumonia CXR images, and 1,228 normal CXR images.

The remainder of the paper is organized as follows. The CXR image dataset, the three pretrained CNN models, the transfer learning and deep features extraction approaches, and the performance analyses are presented in Section II. The results and conclusions are provided in Sections III and IV, respectively.

II. MATERIALS AND METHODS

A. The CXR image dataset

Two publicly available CXR image datasets have been employed in this study. The first dataset is called the Cancer Imaging Archive dataset [17] and it includes 1228 COVID-19 CXR images acquired for subjects with COVID-19 infections. The second dataset is called the RSNA Pneumonia Challenge (RSAN) dataset [18] and it includes 6012 pneumonia CXR images acquired for subjects with pneumonia infections and 8851 normal CXR images acquired for subjects that do not have pneumonia or COVID-19 infections. To avoid the problem of imbalanced data across the three classes, we have randomly selected 1,228 pneumonia CXR images and 1,228 normal CXR images from the RSAN dataset. Hence, the CXR image dataset employed in the current study includes 1,228 COVID-19 CXR images, 1,228 pneumonia CXR images, and 1,228 normal CXR images. Figure 1 shows sample COVID-19, pneumonia, and normal CXR images obtained from the dataset used in the current study.

B. The AlexNet, VGG19, and ResNet50 CNN models

Three CNN models are utilized to classify the CXR images, where these models are the AlexNet model [14], the VGG19 model [15], and the ResNet50 model [16]. The architectures of the AlexNet, VGG19, and ResNet50 models are provided in Fig. 2. As shown in Fig. 2(a), the AlexNet model includes 5 convolution layers and 3 fully connected layers. The architecture of the VGG19 model, which is presented in Fig. 2(b),

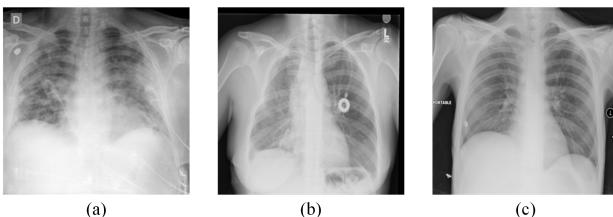


Fig. 1: Sample (a) COVID-19, (b) pneumonia, and (c) normal CXR images.

includes 16 convolution layers and 3 fully connected layers. The ResNet50 model, which is shown in Fig. 2(c), includes 33 convolution layers and 1 fully connected layer, without any hidden fully connected layers. The ResNet50 model also includes shortcut residual connections, which can jump over one or more layers.

The three CNN models are pretrained using the ImageNet database [19], which is composed of millions of natural images grouped into 1000 classes. Therefore, the last fully connected layer of the AlexNet, VGG19, and ResNet50 models is composed of 1000 neurons, which match the 1000 image classes included in the ImageNet database.

C. Classifying the CXR images using the transfer learning approach

The transfer learning approach is the first approach that is employed to classify the CXR images using the ImageNet-pretrained Alexnet, VGG19, and ResNet50 CNN models. In this approach, each ImageNet-pretrained CNN model is reconfigured and fine-tuned to classify the CXR images as COVID-19, pneumonia, or normal based on the procedure described in [20]. In particular, the last fully connected layer of the ImageNet-pretrained CNN model is reconstructed to have 3 neurons that match the three classes of CXR images. In addition, the last fully connected layer is initialized randomly. The CXR images are utilized to fine-tune the reconfigured CNN model. The learning rates employed during the fine-tuning process are equal to 0.01 for the last fully connected layer and 0.001 for all other layers. Furthermore, the stochastic gradient descent method [21] is used to run the fine-tuning process for 50 epochs.

D. Classifying the CXR images using the deep features extraction approach

The deep features extraction approach is the second approach that we have employed to classify the CXR images using the ImageNet-pretrained AlexNet, VGG19, and ResNet50 CNN models. The process of extracting the deep features is carried out by applying the CXR images to each one of the three CNN models and applying the deep features extraction procedure described in [22].

For the AlexNet model, the features maps of the convolution layers CONV3, CONV4, and CONV5, which are shown in Fig. 2(a), are extracted to obtain 1,024 feature maps. The maximum value and the average value of each extracted feature map are computed and used as deep features. Furthermore, the activations of the fully connected layers FC6 and FC7, which are shown in Fig. 2(a), are extracted and employed as deep features maps. Hence, the AlexNet model is used to extract 9,216 deep features (4,096 features from FC6, 4,096 features from FC7, and 2,048 features from CONV3, CONV4, and CONV5).

For the VGG19 model, the features maps of the convolution layers CONV5_1, CONV5_2, CONV5_3, and CONV5_4, which are shown in Fig. 2(b), are extracted to obtain 2,048 feature maps. The maximum value and the average value of

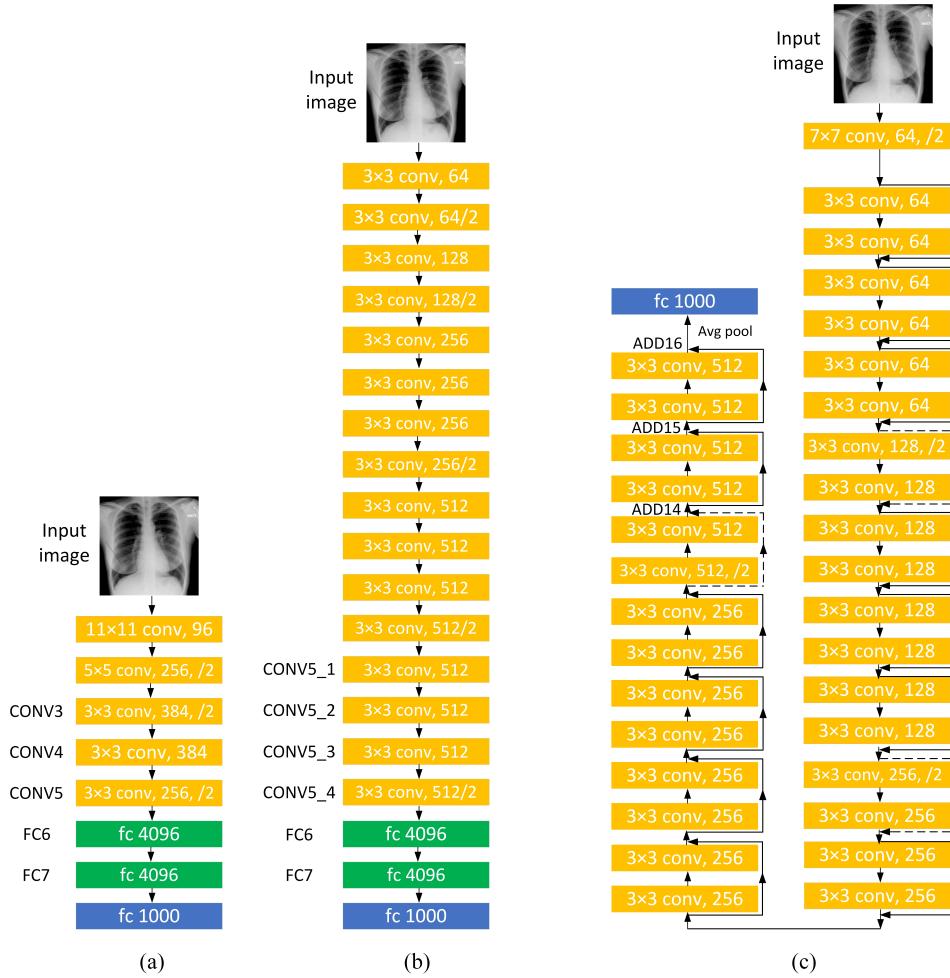


Fig. 2: The architectures of the (a) AlexNet, (b) VGG19, and (c) ResNet50 CNN models.

each extracted feature map are computed and used as deep features. In addition, the activations of the fully connected layers FC6 and FC7, which are shown in Fig. 2(b), are extracted and used as deep features. Hence, the VGG19 model is used to extract 12,288 deep features (4,096 features from FC6, 4,096 features from FC7, and 4,096 features from CONV5_1, CONV5_2, CONV5_3, and CONV5_4).

For the ResNet50 model, the features maps of the addition layers ADD14, ADD15, and ADD16, which are shown in Fig. 2(c), are extracted to obtain 6,144 feature maps. The maximum value and the average value of each extracted feature map are computed and used as deep features. Hence, the ResNet50 model is used to extract 12,288 deep features.

The deep features extracted from each ImageNet-pretrained CNN model might have redundant and irrelevant classification data. Hence, the deep features extracted from each CNN model are analyzed using the two-phase features selection procedure presented in [22], [23]. Assume that P features have been extracted from a given CNN model. The first phase of the features selection procedure ranks the P features using the Minimum Redundancy Maximum Relevance algorithm [24]. The ranked features are processed to generate P feature

groups, such that the p^{th} group includes the top-ranked p features, where $p = \{1, 2, 3, \dots, P\}$. The classification accuracy of each one of the P feature groups is evaluated. The smallest feature group that maximizes the classification accuracy is selected and used as candidate selected features. In the second phase, a backward elimination procedure is utilized to process the candidate selected features. Let Q be the number of the candidate selected features. The first iteration of the backward elimination procedure performs a single feature elimination process to find the $Q - 1$ features that maximize the classification accuracy. To conduct the single feature elimination process, each feature in the Q candidate selected features is temporally removed, and the classification accuracy achieved using the remaining $Q - 1$ features is computed. The $Q - 1$ features that produce the largest enhancement in the classification accuracy are determined and employed as refined selected features. The successive iterations of the backward elimination procedure apply the single feature elimination process in an iterative manner to reduce the size of the selected features and enhance the classification accuracy. The backward elimination procedure stops when the removal of any additional feature leads to a decrease in the classification

accuracy.

For each ImageNet-pretrained CNN model, the selected deep features are analyzed using a Support Vectors Machine (SVM) [12] classifier with the radial basis function kernel. The classifier is fine-tuned as described in [22]. The classification is configured to classify the CXR images using the selected deep features into three classes, namely COVID-19, pneumonia, and normal.

E. Performance evaluation

A tend-fold cross validation procedure in used to classify the CXR images using the three CNN models based on the transfer learning approach and the deep features extraction approach. In this cross validation procedure, the CXR image dataset is randomly divided into ten subgroups. Each one of the first nine subgroups includes 122 COVID-19 CXR images, 122 pneumonia CXR images, and 122 normal CXR images. Moreover, the tenth subgroup includes 130 COVID-19 CXR images, 130 pneumonia CXR images, and 130 normal CXR images. In each fold of the cross-validation procedure, nine CXR image subgroups are used as training images, and the remaining CXR image subgroup is employed as testing images. The cross-validation procedure is run for 10 folds to ensure that each CXR image subgroup is employed as testing images. For the transfer learning approach, the training process aims to fine-tune each one of the CNN models. For the deep features extraction approach, the training process aims to tune and train the SVM classifier.

The classification performance achieved using the six combinations of the three CNN models (AlexNet, VGG19, and ResNet50) and the two classification approaches (the transfer learning and deep features extraction approaches) was measured. Three metrics have been utilized to assess the classification performance, where these metrics are the sensitivity, specificity, and F1 score [25]. The three metrics are computed for each class of CXR images. Furthermore, the macro-averaged values of the three metrics are calculated across the three classes of CXR images.

III. RESULTS

Figure 3 shows the confusion matrices achieved based on the transfer learning approach using the pretrained AlexNet model, VGG19 model, and ResNet50 model. Furthermore, Table I presents the sensitivity, specificity, and F1 score values achieved based on the transfer learning approach using the three CNN models. The ResNet50 model achieved the highest classification performance based on the transfer learning approach. Particularly, the ResNet50 model obtained sensitivity, specificity, and F1 score values equal to 95.7%, 99.3%, and 97.1%, respectively, for the COVID-19 class, 88.3%, 96.5%, and 90.4%, respectively, for the pneumonia class, and 94.9%, 93.6%, and 91.4%, respectively, for the normal class. Additionally, the ResNet50 model achieved macro-averaged sensitivity, specificity, and F1 score values equal to 92.9%, 96.5%, and 93.0%, respectively.

Figure 4 presents the confusion matrices obtained based on the deep features extraction approach using the pretrained AlexNet model, VGG19 model, and ResNet50 model. Moreover, Table II shows the classification sensitivity, specificity, and F1 score values achieved based on the deep features extraction approach using the three CNN models. The results show that the ResNet50 model obtained the highest classification performance based on the deep features extraction approach. Particularly, the ResNet50 model achieved sensitivity, specificity, and F1 score values equal to 96.1%, 99.3%, and 97.4%, respectively, for the COVID-19 class, 89.4%, 96.9%, and 91.5%, respectively, for the pneumonia class, and 95.7%, 94.3%, and 92.4%, respectively, for the normal class. Furthermore, the ResNet50 model obtained macro-averaged sensitivity, specificity, and F1 score values equal to 93.7%, 96.9%, and 93.7%, respectively.

The results presented in Tables I and II indicate that for the three CNN models, the deep features extraction approach obtained higher classification performance compared with the transfer learning approach. The results also show that the highest classification performance reported across the three CNN models and the two classification approaches is achieved by applying the deep features extracted to the ResNet50 model. However, the difference between the classification results obtained based on the deep features extraction approach and the results achieved based on the transfer learning approach is small. This can be attributed to the fact that the fine tuning process performed in the transfer learning approach is carried out using relatively large and balanced images, *i.e.* 1,228 images from each class. Hence, the transfer learning approach enabled the three CNN models to achieve effective differentiation between the three classes of chest X-ray images.

IV. CONCLUSION

Three ImageNet-pretrained CNN models, namely the AlexNet, VGG19, and ResNet50 models, are employed to classify CXR images based on the transfer learning and deep features extraction approaches. The image dataset utilized in our study includes 1,228 COVID-19 CXR images, 1,228 pneumonia CXR images, and 1,228 normal CXR images. In the transfer learning approach, each ImageNet-pretrained CNN model is fine-tuned to classify the CXR images as COVID19, pneumonia, and normal. In the deep features extraction approach, each ImageNet-pretrained CNN model is employed to extract deep features for characterizing the CXR images. Moreover, the deep features extracted using each CNN model are classified using a SVM classifier. The results indicate that the ResNet50 model achieved the highest classification results using the two classification approaches. Additionally, for the three CNN models, the deep features extraction approach achieved better classification results compared with the transfer learning approach. The highest macro-averaged sensitivity, specificity, and F1 score values, which are achieved by applying the deep features extraction approach to the ResNet50 CNN model, are equal to 93.7%, 96.9%, and 93.7%, respectively.

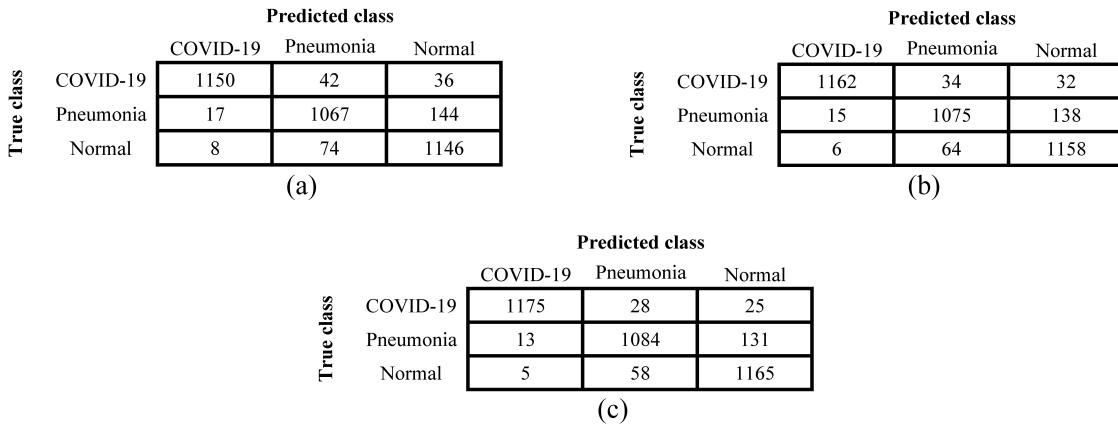


Fig. 3: The confusion matrices obtained based on the transfer learning approach using the (a) AlexNet CNN model, (b) VGG19 CNN model, and (c) ResNet50 CNN model.

TABLE I: The sensitivity, specificity, and F1 score that are obtained based on the transfer learning approach using the (a) AlexNet CNN model, (b) VGG19 CNN model, and (c) ResNet50 CNN model.

CNN model	Class	Sensitivity	Specificity	F1 Score
AlexNet CNN model	COVID-19	93.6%	99.0%	95.7%
	Pneumonia	86.9%	95.3%	88.5%
	Normal	93.3%	92.7%	89.7%
	Macro-averaged values	91.3%	95.6%	91.3%
VGG19 CNN model	COVID-19	94.6%	99.1%	96.4%
	Pneumonia	87.5%	96.0%	89.5%
	Normal	94.3%	93.1%	90.6%
	Macro-averaged values	92.2%	96.1%	92.2%
ResNet50 CNN model	COVID-19	95.7%	99.3%	97.1%
	Pneumonia	88.3%	96.5%	90.4%
	Normal	94.9%	93.6%	91.4%
	Macro-averaged values	92.9%	96.5%	93.0%

REFERENCES

- [1] (2022) COVID-19 Coronavirus Pandemic. [Online]. Available: <https://www.worldometers.info/coronavirus/>
- [2] A. Garg, U. Ghoshal, S. S. Patel, D. V. Singh, A. K. Arya, S. Vasanth, A. Pandey, and N. Srivastava, "Evaluation of seven commercial RT-PCR kits for COVID-19 testing in pooled clinical specimens," *Journal of Medical Virology*, vol. 93, pp. 2281–2286, 2021.
- [3] X. Xie, Z. Zhong, W. Zhao, C. Zheng, F. Wang, and J. Liu, "Chest CT for typical coronavirus disease 2019 (COVID-19) pneumonia: Relationship to negative RT-PCR testing," *Radiology*, vol. 296, pp. E41–E45, 2020.
- [4] L. A. Rousan, E. Elobeid, M. Karrar, and Y. Khader, "Chest x-ray findings and temporal lung changes in patients with COVID-19 pneumonia," *BMC Pulmonary Medicine*, vol. 20, 2020, Article ID 245.
- [5] A. Abbas, M. M. Abdelsamea, and M. M. Gaber, "Classification of COVID-19 in chest x-ray images using DeTraC deep convolutional neural network," *Applied Intelligence*, vol. 51, pp. 854–864, 2021.
- [6] H. S. Alghamdi, G. Amoudi, S. Elhag, K. Saeedi, and J. Nasser, "Deep learning approaches for detecting COVID-19 from chest X-ray images: a survey," *IEEE Access*, vol. 9, pp. 20235–20254, 2021.
- [7] S. R. Nayak, D. R. Nayakb *et al.*, "Application of deep learning techniques for detection of COVID-19 cases using chest X-ray images: A comprehensive study," *Biomedical Signal Processing and Control*, vol. 64, 2021, Article ID 102365.
- [8] M. M. Rahaman, C. Li, Y. Yao, F. Kulwa, M. A. Rahman, Q. Wang, S. Qi, F. Kong, X. Zhu, and X. Zhao, "Identification of COVID-19 samples from chest X-ray images using deep learning: A comparison of transfer learning approaches," *Journal of X-Ray Science and Technology*, vol. 28, pp. 821–839, 2020.
- [9] A. Rehman, S. Naz, A. Khan, A. Zaib, and I. Razzak, "Improving coronavirus (COVID-19) diagnosis using deep transfer learning," *medRxiv*, pp. 1–16, 2020.
- [10] Y. A. M.I. Daoud *et al.*, "COVID-19 diagnosis in chest X-ray images by combining pre-trained CNN models with flat and hierarchical classification approaches," in *Proc. 12th International Conference on Information and Communication Systems (ICICS)*, 2021, pp. 330–335.
- [11] M. Turkoglu, "COVIDetectioNet: COVID-19 diagnosis system based on x-ray images using features selected from pre-learned deep features ensemble," *Applied Intelligence*, vol. 51, pp. 1213–1226, 2021.
- [12] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, United States: Springer, 2000.
- [13] C. Sitaula and S. Aryal, "New bag of deep visual words based features to classify chest x-ray images for COVID-19 diagnosis," *Health Information Science and Systems*, vol. 9, 2021, Article ID 24.
- [14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. International Conference on Neural Information Processing Systems*, 2012, pp. 1097–1105.
- [15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd International Conference*

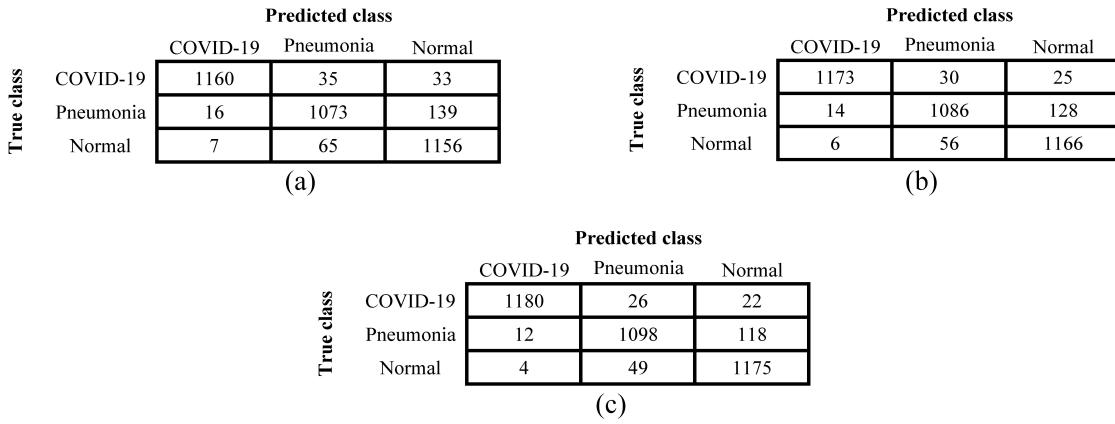


Fig. 4: The confusion matrices obtained based on the deep features extraction approach using the (a) AlexNet CNN model, (b) VGG19 CNN model, and (c) ResNet50 CNN model.

TABLE II: The sensitivity, specificity, and F1 score that are obtained based on the deep features extraction approach using the (a) AlexNet CNN model, (b) VGG19 CNN model, and (c) ResNet50 CNN model.

CNN model	Class	Sensitivity	Specificity	F1 Score
AlexNet CNN model	COVID-19	94.5%	99.1%	96.2%
	Pneumonia	87.4%	95.9%	89.4%
	Normal	94.1%	93.0%	90.5%
	Macro-averaged values	92.0%	96.0%	92.0%
VGG19 CNN model	COVID-19	95.5%	99.2%	96.9%
	Pneumonia	88.4%	96.5%	90.5%
	Normal	95.0%	93.8%	91.6%
	Macro-averaged values	93.0%	96.5%	93.0%
ResNet50 CNN model	COVID-19	96.1%	99.3%	97.4%
	Pneumonia	89.4%	96.9%	91.5%
	Normal	95.7%	94.3%	92.4%
	Macro-averaged values	93.7%	96.9%	93.7%

- on Learning Representations (ICLR2015), 2015.
- [16] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
 - [17] K. Clark, B. Vendt *et al.*, “The cancer imaging archive (TCIA): Maintaining and operating a public information repository,” *Journal of Digital Imaging*, vol. 26, pp. 1045–1057, 2013.
 - [18] “RSNA Pneumonia Detection Challenge,” <https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data>, accessed: June 2021.
 - [19] O. Russakovsky, J. Deng *et al.*, “ImageNet large scale visual recognition challenge,” *Int. J. Comput. Vis.*, vol. 115, pp. 211–252, 2015.
 - [20] H. C. Shin, H. R. Roth *et al.*, “Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning,” *IEEE Trans. Med. Imaging*, vol. 35, pp. 1285–1298, 2016.
 - [21] C. De Sa, M. Feldman, C. Ré, and K. Olukotun, “Understanding and optimizing asynchronous low-precision stochastic gradient descent,” in Proc. 44th Annual International Symposium on Computer Architecture, 2017, pp. 561–574.
 - [22] M. Daoud, S. Abdel-Rahman *et al.*, “Breast tumor classification in ultrasound images using combined deep and handcrafted features,” *Sensors*, vol. 20, 2020, Article ID 6838.
 - [23] M. I. Daoud, S. Abdel-Rahman, and R. Alazrai, “Breast ultrasound image classification using a pre-trained convolutional neural network,” in Proc. 15th International Conference on Signal-Image Technology Internet-Based Systems, 2016, pp. 167–171.
 - [24] H. Peng, F. Long, and C. Ding, “Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 1226–1238, 2005.
 - [25] C. Nicholson, “Evaluation metrics for machine learning - accuracy, precision, recall, and F1 defined,” <https://wiki.pathmind.com/accuracy-precision-recall-f1>.