Analysis of COVID-19 Detection Algorithms Based on Convolutional Neural Network Models Using Chest X-ray Images

Archana R. Nair[(✉)] and A. S. Remya Ajai[(✉)]

Department of Electronics and Communication Engineering, Amrita Vishwa Vidyapeetham, Amritapuri, India
archanarn5974@gmail.com, remya@am.amrita.edu

Abstract. COVID-19, a disease caused by corona virus is a worldwide pandemic which put millions into death. Not only on lives of people but also it has affected all countries in terms of development and wealth. The main challenge is to detect COVID-19 effectively with high accuracy. A fast classification algorithm can help the health professionals in many ways. This work focuses on the implementation analysis of various Convolutional Neural Network (CNN) which are pre-trained in detecting the disease from chest X-ray images with highest accuracy. Transfer learning is utilized, and fine tuning is performed to get a reliable classification of the image data. For the pre-trained CNN model Mobilenet-V2, highest accuracy of 94.298% and precision of 89.76% obtained.

Keywords: COVID-19 · CNN · Transfer learning · Classification · Fine tuning · Performance metrics

1 Introduction

The COVID-19 disease is treated as a global pandemic. On 31st December 2019, World Health Organization (WHO) gave an alert on a special case of respiratory illness with symptoms like fever and cough. By January 30, 2020 WHO promulgated it as a serious global health emergency and gave the name to the disease as COVID-19 [1]. It snatched millions of lives and still countries are suffering from this pandemic. Early detection of disease is the only way to reduce the spread [2, 3].

RTPCR (Reverse Transcription-Polymerase Chain Reaction) and antigen test are proved to be efficient in detecting COVID-19 [4]. But insufficient test kits with exponentially increasing patients compelled health professionals to look for an alternative. Today, detection of disease with radiology images is getting more attraction [5]. Although CT scans are more effective in detection, the increasing number of patients and extra burden on radiologists caused to move from CT scans to Chest X-rays [6].

After COVID-19 outbreak, researchers began to explore the possibilities of deep learning to detect and classify the disease. Machine learning is considered as a subset of Artificial Intelligence (AI) containing approaches to improve the capability of machines

in doing various tasks with experience. It gives the ability for computers to learn and perform. Using Machine Learning approach, patterns from large, noisy and complex datasets can be extracted. This feature of machine learning is what makes it highly suitable for biomedical detection and classification applications [7–9]. Artificial Neural Network (ANN) constitutes a major area under machine learning. It is a brain inspired technique which uses artificial neurons which mimics the actual human brain neurons. Deep learning is a part of ANN which gains major attraction these days. Classical approaches for biomedical image classification consist of hand-crafted feature extraction and then recognition [10]. But in popular deep learning models like Convolutional Neural Network (CNN), manually doing feature extraction is not needed.

Another neural network gaining more focus these days is Spiking Neural Network (SNN) [11]. The important feature of SNN is that it closely mimics a biological neuron. The problem in using SNN is that it lacks a learning method developed specifically for training. Also, it is computationally intensive. For image classification purposes, there is a need to transform the images into spikes. It is found that it will be a lossy process. Also, a good performance cannot be expected. Because of all these reasons, CNN is preferred over SNN for this project.

A pretrained model approach is employed in this paper. For training a neural network from beginning, large dataset is required. In the case of COVID-19 classification, getting a huge dataset is practically impossible. Transfer learning is the method to use a pretrained neural network. Training time is largely reduced by using this approach. It serves as a better initial model and offers faster training rate. With a good initial point and elevated learning rate, the machine learning model converge at a higher performance level. It results in highly accurate output after training.

Here seven pretrained networks are selected for analysis which exhibited good performance as given by various literatures. They are Googlenet, Alexnet, Squeezenet, Mobilenetv2, Resnet-50, Resnet-101 and Vgg19. The mentioned neural networks are analysed based on different performance parameters like accuracy, specificity, precision, sensitivity, and f-score. All implementation and analysis were based on Kaggle dataset. There were two categories of images: normal and COVID-19 affected chest X-ray images. The task is to accurately predict whether a chest X-ray image fed to the model is of a COVID-19 affected person or that of a normal man.

In the pretrained network, fine tuning is performed in the last layers to suit the current classification task. Fine tuning upgrades the accuracy of the model by a large margin. A technique called data augmentation is also employed to combat non availability of huge dataset. Data augmentation includes flipping, rotation and scaling of images.

Detailed literature review is given in Sect. 2. Section 3 describes the methodology adopted and about performance parameters used in this work. Results obtained and its discussion is presented in Sect. 4 and the conclusion is given in Sect. 5.

2   Literature Review

Sharma et al. put forwards the performance assessment of various CNN architectures for object recognition from real time video such as Alexnet, Googlenet and ResNet50. The analysis shows that Googlenet and ResNet50 showed advancement in performance while

comparing with Alexnet. CIFAR 10 and CIFAR 100 datasets were used for this project and only prediction accuracy is considered as performance metric. With CIFAR 100, the model obtained accuracy as follows. For AlexNet: 44.10%, GoogleNet: 64.40%, and ResNet 50: 59.82%. By using CIFAR 10, the accuracy values obtained were, AlexNet: 36.12%, GoogleNet: 71.67%, ResNet 50: 78.10% [12].

Maeda-Gutiérrez et al. did the work to classify diseases caused to tomato plant into ten classes. This work focuses on fine tuning of CNN model. Here, analysis is done using AlexNet, GoogleNet, Inception V3, ResNet 18, and ResNet 50. Nine distinct classes of tomato plant diseases and another healthy class from PlantVillage are used as dataset. This dataset is created by the owners of the paper itself by directly taking images from farm. This work used so many performance metrics for evaluation, like accuracy, precision, sensitivity, specificity, F-score, Area Under the Curve (AUC), and time. Work done is mainly concentrated on fine tuning, that is performed in the last 3 layers of pre trained models to improve performance. The inferences obtained from this paper are as follows. Performance of GoogleNet is better when compared with other similar networks. Inception V3 is one of the deepest CNN architectures. But it shows poor performance [13].

Analysis on six classifiers like: Alexnet, Googlenet, VGG-16, Resnet, InceptionResNet-V2 and Darknet 19 were carried out by Benali et al. In the analysis, it is found that Inception-ResNet-V2 network outperformed all other nets while considering Top-1 and Top-5 accuracy which are 80.3% and 95.1% respectively. Inferred that, architectures rooted in residual concept attain elevated accuracy because of negligible number of parameters while comparing with other architectures. Alexnet underperformed in terms of accuracy and number of MACs are very less compared to Inception-Resnet-V2. Best performers in object detection application are found to be VGG and ResNets [14].

Xiao et al. did a work based on detecting the masks wore by workers. VGG-19 is used in this project. It is proposed that three fully connected layers of this network should be restored with one Flatten layer and two fully connected layers. The advanced version of training the model mainly used the fine-tuning method in transfer learning. Using this technique, parameters of pre-trained VGG-19 CNN model is transferred to the convolution layer, pooling layer and fully connected layer of the proposed model to detect masks [15].

Demir et al. seek to find application of their work in biomedical field. Here, for detecting cancer at initial stages, CNN models like Inception-V3 and ResNet101 are implemented with the help of datasets containing skin cancer images. Dataset is formed with 2437 images having size $224 \times 224 \times 3$. It is composed of malignant and benign images which is taken from ISIC-Archive database. For ResNet101 and Inception-V3, the results from this work stated an accuracy of 84.09% and 87.42% respectively [16].

Lungs disease classification can be an application of CNN and is explained by Shin et al. In this paper, for identifying the interstitial lung disease, 5 distinct methods of neural networks are presented which are CNN based. Dataset consists of CT scan slices (2D). It carries images with a count of 905 of 120 people having 6 different kind of lung tissue conditions. Model is trained with the help of architectures like CifarNet, GoogleNet, AlexNet, and ImageNet. Also implementation of various architectures is done to detect the Thoracoabdominal Lymph node and made keen evaluation [17].

3   Methodology

The workflow is presented as a block diagram and is given in Fig. 1.
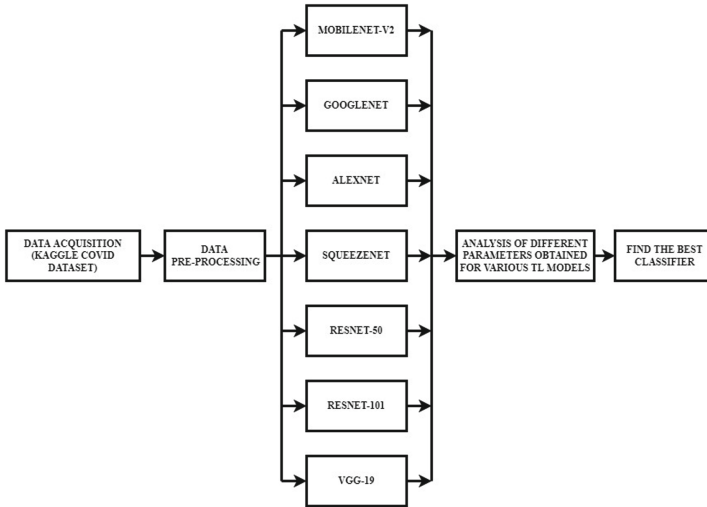

Fig. 1. Basic block diagram showing workflow

COVID-19 dataset is obtained from Kaggle database [18, 19]. Only images satisfying a minimum quality is selected for training the model. Then the images are subjected to the training phase. For that, seven renowned classifiers are selected which include Mobilenet-v2, Googlenet, Alexnet, Squeezenet, Resnet-50, Resnet-101 and VGG-19. The task is to find which classifier is best to classify COVID-19. To find this, several performance metrics are used. As it's a biomedical application, more focus will be given to test accuracy and sensitivity.

3.1   Data Preprocessing

This work focuses on COVID-19 disease classification. The dataset count is 3616. It consists of chest X-ray images of patients affected with COVID-19 and 10,000 images of normal chest X-rays. Sample images for each category is given as Fig. 2. The image count chosen for training and testing purpose is shown in Table 1.
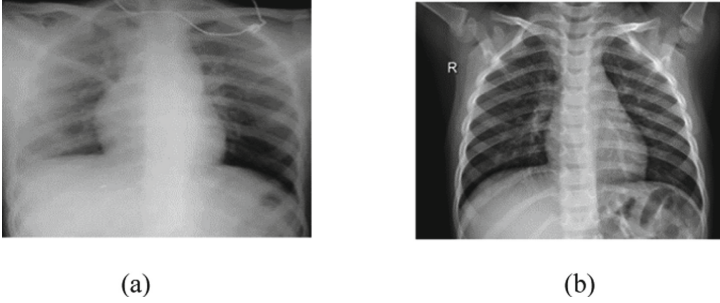
(a)   (b)

Fig. 2. (a) Chest X-ray of a COVID-19 affected person. (b) Chest X-ray of a normal person

Table 1. Dataset

	Image count	
	COVID-19	Normal
Training set and validation set	326	326
Testing set	100	100

It is clear from the above table, the image count in each category is balanced. Entire images contained in dataset are in Portable Network Graphics(png) file format. 1024 × 1024 pixels or 256 × 256 pixels are the resolution of images. All the pretrained models selected need images of size 224 × 224 × 3 pixels. So, the images are resized prior to feeding into the model. Training and validation images are selected randomly with a 7:3 ratio respectively. Data augmentation is the method used over the images to raise the dataset count. Data augmentation includes flipping, rotation and scaling of images.

3.2 Transfer Learning

Transfer learning can be viewed as a technique of reutilizing a pre-trained model for a new piece of work [20]. It saves training time and improves accuracy. In this paper, seven pretrained models are analysed which includes Googlenet, Alexnet, Squeezenet, Mobilenetv2, Resnet-50, Resnet-101 and Vgg19.

Alexnet
Alexnet holds 60 million parameters and 650000 neurons [21]. It consists of eight layers in which five are convolutional layers and next three are fully connected layers. There is max pooling layer and softmax classification layer. Inorder to prevent overfitting, a dropout ratio of 0.5 is also provided. Instead of using saturating activation functions like Tanh or Sigmoid, the authors used a special activation function called Rectified Linear

Unit (ReLU) non-linearity inorder to have a faster training. ReLU is present in the first seven layers of Alexnet.

Googlenet
Unlike Alexnet, Googlenet possess only seven million parameters [22]. It has 4 convolutional layers, 9 inception modules, 4 max pooling and 3 average pooling layers, 5 fully connected layers and 3 softmax classifier layers. It is a wider CNN with 22 layers. It uses a dropout ratio of 0.7 and ReLU activation function in convolutional layers. It can be said that Googlenet is a powerhouse having increased computational efficiency.

Residual Network (ResNet)
ResNet is based on deep architectures, and it uses a technique called skip connections [23]. That is, training from a few layers is skipped and it connects the intermediate layers directly to the output. It upgrades the performance of the architecture to a large extent. The residual unit contains convolutional, max pooling and fully connected layers. ResNet can have 18, 34, 50, 101, 152 and 1202 deep layers. In this project, ResNet-50 and ResNet-101 are selected for analysis. Computing resources and training time rises with rise in the number of deep layers.

Squeezenet
Squeezenet architecture is a smaller CNN architecture which doesn't show compromise in accuracy [24]. It has 50 times fewer parameters than Alexnet while achieving the same accuracy as that of Alexnet. It can be easily implemented on FPGAs having limited memory. The authors were able to compress the model to less than 0.5 MB. The base of Squeezenet architecture is the fire module which is made up of a squeeze layer and an expand layer. The main features include it uses 1×1 filters instead of 3×3 filters and pooling is done at late stage only resulting in large activation maps for convolutional layers.

Mobilenetv2
For mobile and resource constrained environments, a new architecture called Mobilenetv2 was introduced [25]. Here, number of parameters and memory requirement is greatly reduced while maintaining the accuracy. A regular convolution operation will do both filtering and combining in a single go, but in Mobilenetv2 these operations are done as separate steps. It is called depthwise separable convolution. ReLU6 is used as the activation function. This architecture uses very less computation power to run or for transfer learning purposes.

VGG-19
VGG-19 can be considered as the successor of Alexnet. It has 19 layers comprising of sixteen convolutional layers, 3 fully connected layers and a final softmax layer [26]. The model achieved good classification accuracy. This architecture can be used for a variety of complex tasks like face recognition.

In a CNN, the convolutional layers present will extract image features which are low level and high level. Final learnable layer and classification layer will classify the applied feed in image based on the output from convolutional layers. When pretrained

networks are used, these final layers are replaced with new layers which are adapted to new classification task and dataset. Fully connected layer acts as final layer with learnable weights in most of the networks. In the new classification task, this layer is superseded with a novel fully connected layer having outputs equal to the number of classes in new task. Generally learning rate is increased in the new layers while performing transfer learning. Freezing of layers is performed to speed up the network training. Freezing means setting of learning rate in the earlier layers of a network to zero. Freezing of layers also has an advantage that it prevents overfitting if the new dataset is small. Data augmentation is performed for resizing the training images to the desired size as required by the network. Also, data augmentation is used to increase dataset by performing flipping, translating and scaling of images. No learning is there in initial layers as it is frozen. Middle layers having slower learning and final layers equipped with fast learning is preferable to speed up training process. Training options make use of stochastic gradient descent with momentum. The advantage of using this technique is that it converges better with longer training time.

3.3 Performance Parameters

The performance of the network is evaluated by generating the confusion matrix. Confusion matrix gives the information about predicted class and actual class [27].

The efficiency indicators such as accuracy, specificity, precision, sensitivity, and F score are evaluated [13]:

Accuracy
It gives the fraction of predictions that the trained model gives as correct. For e.g., an accuracy of 80% means 2 of every 10 labels is incorrect and remaining 8 are correctly predicted.

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

where TP-True Positive, FN-False Negative, FP-False Positive, TN-True Negative.

Specificity
It gives the number of correct predictions from all X-rays which are actually normal. For e.g., a specificity of 70% means 3 of every 10 normal people are mislabeled as COVID-19 affected and 7 are correctly labelled as normal.

$$Specificity = \frac{TN}{TN + FP}. \tag{2}$$

Precision
It gives an estimation about the proportion of positive identifications which are actually right. For e.g., if precision is found to be 50%, it means when an X-ray is predicted as COVID-19 affected, it is correct 50% of time.

$$Precision = \frac{TP}{TP + FP}. \tag{3}$$

Sensitivity (Recall)

It tells what proportion of actual positives were identified as such. For eg, a sensitivity of 60% means the model correctly identifies 60% of all COVID-19 affected X-rays.

$$Sensitivity = \frac{TP}{TP + FN}.$$

(4)

F score

It considers both precision and sensitivity. Here, harmonic mean of precision and recall is calculated.

$$F\ score = 2 * \frac{Precision * Recall}{Precision + Recall}.$$

(5)

4 Results and Discussion

The required chest X-ray images of COVID-19 affected and healthy were collected from Kaggle dataset. This project is done with the help of MATLAB 2021a. The images were stored in an image datastore. Used 70% of images for training and 30% for validation purpose. The test images were stored as a separate set. As mentioned earlier, seven pre-trained networks were utilized to find the best classifier. All models were trained first and then tested to find the performance parameters. Table 2 showing summarized results for each pre trained network is given below.

Table 2. Result summary

Net name	Test accuracy	Validation accuracy	Specificity	Precision	Sensitivity	F score
Mobilenet-v2	94.298	98.19	88.6	89.76	100	94.61
Googlenet	93.859	98.19	87.72	89.06	100	94.21
Alexnet	92.982	97.59	85.96	87.69	100	93.44
Squeezenet	93.421	98.19	86.84	88.37	100	93.83
Resnet-50	93.859	98.19	87.72	89.06	100	94.21
Resnet-101	93.716	98.19	88.6	89.66	98.84	94.02
VGG-19	93.859	97.59	87.72	89.06	100	94.21

From analyzing above results, it can be found that Mobilenet-v2 is the best classifier for COVID-19 disease classification. For biomedical applications, the important performance metrics are accuracy and sensitivity. Mobilenet-v2 has a test accuracy of 94.298 and sensitivity of 100%. It means all COVID-19 affected X-rays are identified correctly as COVID-19 affected itself. No COVID-19 cases are missed here even though some

normal X-rays were found as COVID-19 affected. Except for Resnet-101, all classi-
fiers gave a sensitivity of 100%. All classifiers showed a training accuracy of 100% as
expected.

The comparison of test accuracy, specificity, precision, sensitivity, and f-score are
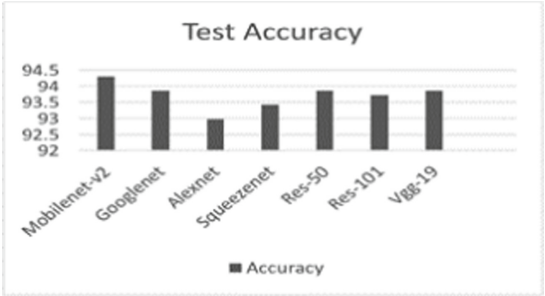presented in Fig. 3, 4, 5, 6 and Fig. 7.

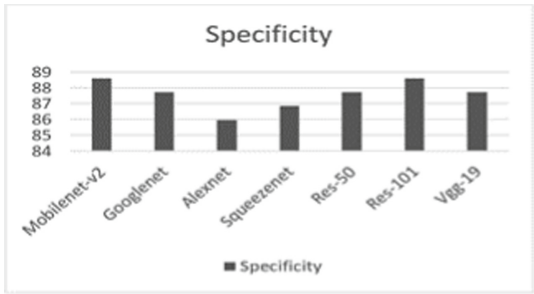**Fig. 3.** Comparison of test accuracy values obtained for considered CNN models.

Fig. 4. Comparison of specificity values obtained for considered CNN models.
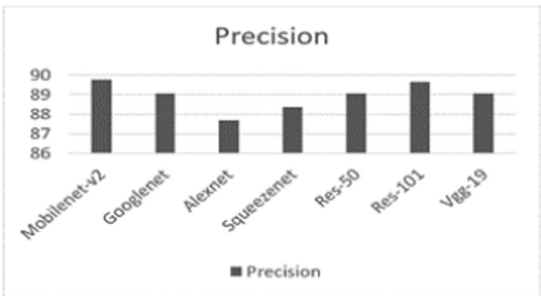

Fig. 5. Comparison of precision values obtained for considered CNN models.

Mobilenet-v2 has fewer parameters (3.5 million) when compared with other pre
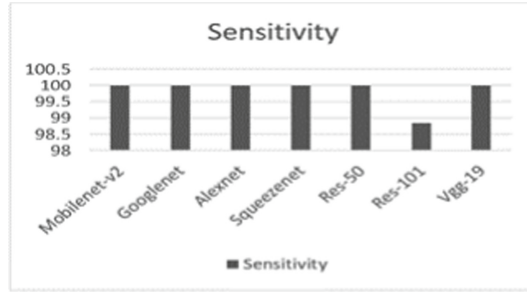trained networks and it leads to faster training and reduced complexity. It also has a

Fig. 6. Comparison of sensitivity values obtained for considered CNN models.
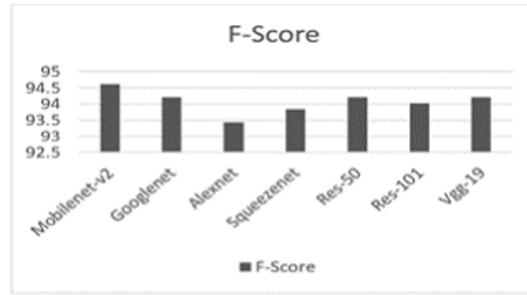

Fig. 7. Comparison of F-score values obtained for considered CNN models.

lower memory requirement of 13 MB. It makes it very suitable for implementation of CNN in a FPGA board.

5   Conclusion

In this paper, an algorithm is implemented to identify the best classifier for classifying COVID-19 disease. Seven classifiers namely Mobilenet-v2, Googlenet, Alexnet, Squeezenet, Resnet-50, Resnet-101 and VGG-19 were studied and implemented in MATLAB 2021a based on their relevance. For analysis purpose, efficiency evaluation metrics like test accuracy, validation accuracy, specificity, precision, sensitivity and f-score for each pre trained network are calculated. All classifiers performed well and thus proves that deep learning networks are promising candidates for complex classification problems. Among seven classifiers, Mobilenet-v2 is found to be the best classifier to classify random chest X-ray images into COVID-19 affected and normal. Utilization of transfer learning and fine tuning of layers are the principal players in this project.

However, the features of images used for training and testing purpose greatly affects the performance of classifiers. In classifying biomedical images, low quality images and the number of images available cause hindrance for building efficient deep learning models. This work can be further broadened to find different lung disorders along with COVID-19. The analysis can also be done with various datasets available where new COVID variants are also considered.

References

1. Chung, M., et al.: CT imaging features of 2019 novel coronavirus (2019-nCoV). Radiology **295**(1), 202–207 (2020). https://doi.org/10.1148/radiol.2020200230

2. Toussie, D., Voutsinas, N., Finkelstein, M., et al.: Clinical and chest radiography features determine patient outcomes in young and middle-aged adults with COVID-19. Radiology **297**(1), E197–E206 (2020)

3. Manikandan, A., Shriram, S., Sarathchandran, C., Palaniappan, S., Rohith, N.D.: Prediction of COVID-19 with supervised regression algorithm through minimum variance unbiased estimator. Int. J. Curr. Res. Rev. **13**(11), s55–s62 (2021)

4. Love, J., et al.: Comparison of antigen- and RT-PCR-based testing strategies for detection of SARS-CoV-2 in two high-exposure settings. PLoS ONE **16**(9), e0253407 (2021). https://doi.org/10.1371/journal.pone.0253407

5. Ghaderzadeh, M., Asadi, F.: Deep learning in the detection and diagnosis of COVID-19 using radiology modalities: a systematic review. J. Healthc. Eng. **2021**, 6677314 (2021). https://doi.org/10.1155/2021/6677314

6. Sun, P., Lu, X., Xu, C., Sun, W., Pan, B.: Understanding of COVID-19 based on current evidence. J. Med. Virol. **92**(6), 548–551 (2020)

7. Cruz, J.A., Wishart, D.S.: Applications of machine learning in cancer prediction and prognosis. Cancer Inform. **2**, 117693510600200030 (2006). https://doi.org/10.1177/117693510600200030

8. Kusuma, S., Divya Udayan, J.: Machine learning and deep learning methods in heart disease (HD) research. Int. J. Pure Appl. Math. **119**(18), 1483–1496 (2018)

9. Remya Ajai, A.S., Gopalan, S.: Analysis of active contours without edge-based segmentation technique for brain tumor classification using SVM and KNN classifiers. In: International Conference on Communication Systems and Networks, ComNet 2019, Thiruvananthapuram, pp. 1–10 (2019)

10. Indumathi, T.V., Sannihith. K., Krishna, S., Remya Ajai, A.S.: Effect of co-occurrence filtering for recognizing abnormality from breast thermograms. In: Proceedings of the Second International Conference on Electronics and Sustainable Communication Systems (ICESC-2021), Coimbatore, pp. 1170–1175 (2021)

11. Vaila, R., Chiasson, J., Saxena, V.: Feature extraction using spiking convolutional neural networks, pp. 1–8, July 2019. https://doi.org/10.1145/3354265.3354279

12. Sharma, N., Jain, V., Mishra, A.: An analysis of convolutional neural networks for image classification. Procedia Comput. Sci. **132**, 377–384 (2018)

13. Maeda-Gutiérrez, V.,et al.: Comparison of convolutional neural network architectures for classification of tomato plant diseases. Appl. Sci. **10**(4), 1245 (2020)

14. Benali Amjoud, A., Amrouch, M.: Convolutional neural networks backbones for object detection. In: El Moataz, A., Mammass, D., Mansouri, A., Nouboud, F. (eds.) ICISP 2020. LNCS, vol. 12119, pp. 282–289. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-51935-3_30

15. Xiao, J., Wang, J., Cao, S., Li, B.: Application of a novel and improved VGG-19 network in the detection of workers wearing masks. J. Phys. Conf. Ser. **1518**(1), 012041 (2020)

16. Demir, A., Yilmaz, F., Kose, O.: Early detection of skin cancer using deep learning architectures: resnet-101 and inception-v3. In: 2019 Medical Technologies Congress (TIPTEKNO), Izmir, Turkey, pp. 1–4 (2019)

17. Shin, H.C., et al.: Deep: convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics, and transfer learning. IEEE Trans. Med. Imaging **35**(5), 1285–1298 (2016)

18. Chowdhury, M.E.H., et al.: Can AI help in screening viral and COVID-19 pneumonia? IEEE Access **8**, 132665–132676 (2020)
19. Rahman, T., et al.: Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images. arXiv preprint arXiv:2012.02238 (2020)
20. Tamil Priya, D., Divya Udayan, J.: Transfer learning techniques for emotion classification on visual features of images in the deep learning network. Int. J. Speech Technol. **23**(2), 361–372 (2020). https://doi.org/10.1007/s10772-020-09707-w
21. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, New York, pp. 1097–1105. Curran Associates, Inc. (2012)
22. Szegedy, C., et al.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015, pp. 1–9 (2015)
23. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016, pp. 770–778 (2016)
24. Iandola, F.N., Han, S., Moskewicz, M.W., Ashraf, K., Dally, W.J., Keutzer, K.: SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size. arXiv:1602.07360 [cs], November 2016
25. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.-C.: MobileNetV2: inverted residuals and linear bottlenecks. arXiv:1801.04381 [cs], March 2019
26. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556 [cs], April 2015
27. Bhagya, T., Anand, K., Kanchana, D.S., Remya Ajai, A.S.: Analysis of image segmentation algorithms for the effective detection of leukemic cells. In: Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019), Tirunelveli, pp. 1232–1236 (2019)