2026 Volume: 02 Issue: 05 www.ijrpa.com ISSN 2456-9995 Research Article

International Journal Research Publication Analysis

Page: 01-13

COVID-19 DETECTION FROM CHEST X-RAY IMAGES USING DEEP LEARNING AND CNN ARCHITECTURES

[*1]Madhu G., [2]Keerthana M. D., [2]Niharika T., [2]Rohith R.

[1]Assistant Professor Department of Computer Science and Engineering, Jyothy Institute of Technology (JIT), Thataguni, Off Kanakapura Road, Bengaluru – 560082, Karnataka, India.

[2]Department of Computer Science and Engineering, Jyothy Institute of Technology (JIT), Thataguni, Off Kanakapura Road, Bengaluru – 560082, Karnataka, India.

Article Received: 27 March 2026

Article Revised: 17 April 2026

Published on: 07 May 2026

***Corresponding Author: Madhu G.**

Assistant Professor Department of Computer Science and Engineering, Jyothy Institute of Technology (JIT), Thataguni, Off Kanakapura Road, Bengaluru – 560082, Karnataka, India. DOI: https://doi-doi.org/101555/ijrpa.5778

ABSTRACT

The COVID-19 pandemic has underscored a pressing demand for fast, precise, and deployable diagnostic solutions capable of supporting clinicians in identifying infected patients. Chest X-ray (CXR) imaging stands out as economical and broadly accessible for pneumonia screening. This study introduces an automated framework for COVID- 19 identification from CXR images, leveraging five CNN architectures—VGG-16, ResNet-50, InceptionV3, DenseNet- 121, and MobileNet-V2—evaluated under transfer learning and scratch training regimes using ImageNet pre-trained weights. Spatial feature enhancement is achieved through histogram equalization and CLAHE preprocessing. Experiments are performed on the COVID-19 Radiography Database and COVIDx dataset, targeting three-class classification: COVID-19, Normal, and Viral Pneumonia. DenseNet-121 with transfer learning recorded the best test accuracy of 97.4% and sensitivity of 96.8%, confirming the value of hierarchical feature reuse in medical imaging contexts.

KEYWORDS: *COVID-19 Detection, Chest X-Ray, Deep Learning, CNN, Transfer Learning, DenseNet-121, ResNet- 50, Medical Image Classification, VGG-16, InceptionV3, CLAHE, Radiography.*

1. INTRODUCTION

The worldwide emergence of COVID-19, originating from the SARS-CoV-2 pathogen, imposed severe strain on global health infrastructure. Early and reliable identification of

infected individuals plays an indispensable role in containing disease propagation and ensuring timely clinical intervention. RT-PCR-based molecular testing, though regarded as the most definitive diagnostic method, carries notable drawbacks including extended turnaround times of 24–72 hours, elevated costs, and false-negative rates ranging from 2% to 29% depending on sampling quality and disease stage [1].

Chest X-ray (CXR) imaging has gained widespread traction in clinical settings due to its rapid acquisition time (< 5 minutes), low cost, and routine availability. COVID-19- specific radiographic findings—ground-glass opacities (GGOs), bilateral peripheral consolidations, and crazy- paving patterns—are detectable in CXR even in early-stage infection, typically appearing within 4–8 days of symptom onset [2]. Deep learning methodologies, especially CNN architectures, have exhibited exceptional capability in automating radiological image interpretation, deriving discriminative spatial representations directly from pixel- level input through hierarchical feature transformations.

This study examines five widely adopted CNN architectures—VGG-16, ResNet-50, InceptionV3, DenseNet-121, and MobileNet-V2—in the context of three- class CXR classification. A structured comparison is conducted across transfer learning and from-scratch training settings, while also assessing preprocessing interventions including histogram equalization and CLAHE, with evaluation using accuracy, sensitivity, specificity, F1-score, and AUC-ROC.

2. MATHEMATICAL FOUNDATIONS

A. Convolutional Operation

A CNN processes an input image of spatial dimensions H × W × C through learnable filters. For a filter F of size k × k applied to input feature map X, the discrete cross- correlation output S is:

$$S(i,j) = \Sigma^{m} \Sigma^{n} X(i+m, j+n)\cdot F(m,n) + b \quad (1)$$

where b is a bias term and m, n index filter positions. Applying ReLU activation over L layers yields hierarchical representations:

$$h^{l} = \mathrm{ReLU}(W^{l} * h^{l-1} + b^{l}), \quad h^{0} = x \quad (2)$$

Max pooling with stride s reduces spatial dimensionality and provides local translation invariance:

$$P(i,j) = \max^{mn} h(si+m, sj+n) \quad (3)$$

B. Residual Learning (ResNet)

He et al. [9] introduced residual shortcut connections to resolve the vanishing gradient problem. Instead of learning H(x) directly, each residual block learns the residual mapping:

$$F(x) := H(x) - x \;\Rightarrow\; H(x) = F(x) + x \quad (4)$$

The forward pass computes $y = F(x, \{W^d\}) + x$. When dimensions differ, a linear projection W_s is applied:

$$y = F(x,\{W^d\}) + W_s x.$$

This enables successful training of networks with 50–152 layers.

C. Dense Connectivity (DenseNet)

DenseNet [11] extends residual connectivity so that each layer l receives feature maps from all preceding layers within a dense block:

$$x^l = H^l([x_0, x_1, \ldots, x^{l-1}]) \quad (5)$$

where $[x_0,\ldots,x^{l-1}]$ denotes channel-wise concatenation and H^l is a composite BN → ReLU → Conv function. A growth rate k controls the number of new feature maps per layer. The total input features to layer l is $k_0 + k(l-1)$, where k_0 is the initial channel count.

D. Depthwise Separable Convolutions (MobileNet-V2)

MobileNet-V2 [12] factorizes standard convolutions into depthwise and pointwise components. A standard convolution with M input channels, N output filters, filter size $D^k \times D^k$ on $D^q \times D^q$ feature maps has cost:

$$Cost_{st} = D^{q2} \cdot M \cdot N \cdot D^{k2} \quad (6)$$

Depthwise separable convolution reduces this by a factor of $1/N + 1/D^{k2}$:

$$Cost^{es} = D^{q2} \cdot M \cdot D^{k2} + D^{q2} \cdot M \cdot N \quad (7)$$

MobileNet-V2 additionally introduces inverted residuals with linear bottlenecks, expanding channels before depthwise convolution and contracting via linear projection to prevent

information loss.

E. Inception Modules (InceptionV3)

Szegedy et al. [10] proposed parallel multi-scale feature extraction, where filters of different sizes operate simultaneously on the same input:

$$\text{Inc}(x) = \text{Concat}[\text{Conv}_{1\times1}(x), \text{Conv}_{3\times3}(x), \text{Conv}_{5\times5}(x), \text{Pool}(x)] \quad (8)$$

InceptionV3 factorizes n×n convolutions into 1×n and n×1 asymmetric sequences, reducing parameter count from 9 to 6 per 3×3 operation (33% reduction).

F. Transfer Learning Formulation

Transfer learning adapts a model pretrained on source domain D_s (ImageNet, 1.28M images, 1,000 classes) to target domain D_t (CXR, 21,165 images, 3 classes). The optimization objective is:

$$\theta_t^* = \text{argmin}_\theta \, L(f(X;\theta), Y_t), \text{ init. from } \theta_s \quad (9)$$

Fine-tuning progressively unfreezes convolutional blocks from top to bottom. The classification head is replaced by: GAP → Dense(256, ReLU) → BatchNorm → Dropout(0.4) → Softmax(3).

G. Loss Function and Optimization

Categorical cross-entropy loss for K = 3 classes over N training samples is defined as:

$$L = -(1/N) \sum^k \sum_i y^k_i \cdot \log(\hat{y}^k_i) \quad (10)$$

where $\hat{y}^k_i = \text{softmax}(z)_i = \exp(z_i) / \sum^d \exp(z^d)$.

The Adam optimizer updates parameters using adaptive moment estimates:

$$m_t = \beta_1 m_{t-1} + (1-\beta_1)\nabla\theta L \quad (11) \qquad v_t = \beta_2 v_{t-1} + (1-\beta_2)(\nabla\theta L)^2 \quad (12) \qquad \theta_{t+1} = \theta_t - \alpha \cdot \hat{m}_t / (\sqrt{\hat{v}_t} + \varepsilon) \quad (13)$$

Hyperparameters:

$\beta_1 = 0.9$, $\beta_2 = 0.999$, $\varepsilon = 10^{-8}$, initial $\alpha = 10^{-4}$

with cosine annealing. Early stopping is applied after 7 patience epochs on validation loss.

H. Evaluation Metrics

From the confusion matrix (TP, TN, FP, FN), the primary classification metrics are computed as follows:

Accuracy = (TP + TN) / (TP + TN + FP + FN) *(14)*

Sensitivity = TP / (TP + FN) *(15)*

Specificity = TN / (TN + FP) *(16)*

Precision = TP / (TP + FP) *(17)*

F1 = 2·Precision·Sensitivity / (Precision + Sensitivity) *(18)*

MCC = (TP·TN − FP·FN) /

√[(TP+FP)(TP+FN)(TN+FP)(TN+FN)] *(19)*

AUC-ROC is computed as the area under the receiver operating characteristic curve: AUC = $\int_0^1$ TPR(t) d(FPR(t)). For multi-class imbalanced settings, macro-averaged F1 is preferred: $F1\_macro = (1/K)\Sigma^k F1^k$.

I. CLAHE Preprocessing

Contrast Limited Adaptive Histogram Equalization (CLAHE) divides the image into non-overlapping tiles and applies histogram equalization locally. For a tile with histogram H(b) and clip limit C_l:

$H^{cl}{}_i{}^p(b) = \min(H(b), C_l)$ *(20)*

$\text{Excess} = \Sigma^b \max(0, H(b)-C_l) / N^b{}_i{}^n{}_s$ *(21)*

Clipped excess pixels are redistributed uniformly across all bins. The equalized output is derived from the cumulative distribution function (CDF):

$I_o{}^u{}_t(x,y) = \text{round}[(CDF(I(x,y))-CDF^m{}_i{}^n) / (N_{til}{}^e-CDF^m{}_i{}^n) \times (L-1)]$ *(22)*

Bilinear interpolation between adjacent tile mappings eliminates boundary artifacts. CLAHE contributed a 1.5– 2.3% sensitivity improvement across all tested architectures by enhancing contrast of ground-glass opacities.

3. EXTENDED LITERATURE SURVEY

A. Early CNN Approaches to COVID-19 CXR Screening

Wang et al. [1] introduced COVID-Net, a purpose-built deep CNN for CXR-based COVID-19 diagnosis, achieving 83.5% test accuracy on the COVIDx benchmark (13,975 CXR

images). COVID-Net employed a human-machine collaborative design methodology guided by generative synthesis optimization to produce a lightweight, COVID- specific architecture. Despite its tailored design, its modest accuracy relative to ImageNet-pretrained models underscores the primacy of transfer learning when target-domain labelled data is limited.

Chowdhury et al. [2] systematically benchmarked VGG- 16, ResNet, and DenseNet variants on the COVID-19 Radiography Database (21,165 images), demonstrating that densely connected networks consistently outperform architectures with fewer skip-connection pathways. This study also curated the standard benchmark dataset used across the field and reported that DenseNet models benefit most from ImageNet initialization due to their dense feature- reuse structure.

Apostolopoulos and Mpesiana [3] evaluated MobileNet, InceptionV3, and VGG-19 across COVID-19, community- acquired pneumonia, and normal CXR classes, achieving 94–96.78% accuracy on 1,427 merged images. This was among the earliest studies demonstrating lightweight architecture viability for mobile-constrained screening.

Sethy and Behera [4] proposed a hybrid strategy combining ResNet-50 feature extraction with an SVM classifier, yielding 95.38% on a 50-image dataset. While architecturally innovative, the extremely small dataset fundamentally limits generalizability—a common limitation of early pandemic studies.

Ozturk et al. [13] developed DarkCovidNet based on the Darknet architecture for binary and multi-class COVID-19 detection from CXR, achieving 98.08% binary accuracy and 87.02% for three-class classification. Their work demonstrated the advantage of purpose-designed lightweight architectures for rapid clinical deployment.

B. Transfer Learning Strategies in Medical Imaging

Tajbakhsh et al. [5] conducted a landmark comparative study across four medical tasks—polyp detection, lung nodule classification, pulmonary embolism detection, and cardiac segmentation—demonstrating that fine-tuning ImageNet-pretrained CNNs reliably surpasses random initialization even when source and target domains differ substantially. Their findings established that low-level CNN features (edges, gradients, textures) are universally applicable across natural and medical imaging domains, while high-level features require task-specific adaptation.

Raghu et al. [6] (Transfusion) demonstrated through centered kernel alignment (CKA) analysis that CNNs pretrained on natural images achieve medical imaging performance comparable to task-specific networks when target dataset sizes are small (< 10,000 samples).

Lower convolutional layers exhibit high representational similarity between ImageNet-pretrained and medical-task-trained networks, justifying progressive unfreezing strategies.

Pan and Yang [14] provided a formal taxonomy of transfer learning, distinguishing inductive ($T_s \neq T_t$), transductive ($D_s \neq D_t$, $T_s = T_t$), and unsupervised transfer settings. COVID-19 CXR classification from ImageNet pretraining falls into the inductive transfer category, where the source and target tasks differ but share spatial feature representations.

C. Preprocessing and Augmentation Strategies

Narin et al. [7] documented that minimal preprocessing pipelines (resize + normalize) led to 2–3% accuracy degradation compared to contrast-enhanced inputs on small COVID-19 datasets. Standard augmentation operations applied in COVID-19 imaging models include horizontal flipping ($p = 0.5$), angular rotation (±15°), zoom perturbation (±10%), brightness adjustment (±20%), and translational shift (±0.1 fraction).

Islam et al. [15] demonstrated that lung-segmented inputs improved COVID-19 detection accuracy by 1.8–2.5% compared to full CXR inputs by reducing the influence of non-pulmonary features such as cardiac silhouette and chest wall. Their combined CNN-LSTM network further exploited temporal radiographic patterns.

D. Hybrid and Ensemble Approaches

Beyond single-architecture studies, hybrid pipelines have demonstrated competitive performance. Zhang et al. [16] combined CNN embeddings with anomaly detection using confidence scores, achieving 96.0% COVID-19-class sensitivity with a controlled false-positive rate. Minaee et al. [17] evaluated ResNet18, ResNet50, SqueezeNet, and DenseNet121 on 5,000 CXR images, reporting that ensemble averaging improved sensitivity by 1.3% compared to the best single model, with DenseNet-121 confirmed as the strongest individual performer.

E. Explainability and Clinical Interpretability

A critical limitation of accuracy-optimized deep learning in clinical settings is the lack of interpretability. Selvaraju et al. [18] introduced Gradient-weighted Class Activation Mapping (Grad-CAM), which generates class-discriminative localization maps by computing gradient-based importance weights α^{kk} for each feature map A^k:

$$L^c_s = \mathrm{ReLU}(\Sigma^k \alpha^{kc} \cdot A^k(x,y)) \quad (23)$$

where $\alpha^{kc} = (1/Z)\Sigma_i\Sigma^j (\partial y^c / \partial A^{k}_{i}{}^{j})$. Grad-CAM overlays on CXR images enable radiologists to verify that model attention aligns with clinically relevant pulmonary regions (lower lobes, peripheral zones), providing explainability essential for regulatory compliance.

4. METHODOLOGY

A. Dataset Description

The experimental dataset used in this work is the COVID-19 Radiography Database assembled by Chowdhury et al. [2] and distributed via the Kaggle platform. It comprises 21,165 CXR images spanning three diagnostic categories: COVID-19 (3,616 images), Normal (10,192 images), and Viral Pneumonia (1,345 images). All images are stored in PNG format at 299×299 pixel resolution. Data partitioning follows an 80:10:10 ratio for training, validation, and test sets. Class imbalance is managed through a combination of weighted sampling and targeted data augmentation.

B. Preprocessing Pipeline

A uniform preprocessing sequence is applied to all input images: (i) spatial resizing to match each architecture's expected input (224×224 for VGG-16, ResNet-50, DenseNet-121, and MobileNet-V2; 299×299 for InceptionV3); (ii) local contrast enhancement via CLAHE (equations 20–22); (iii) pixel intensity normalization to [0, 1]; and (iv) channel-wise mean subtraction based on ImageNet statistics during transfer learning. Training-time augmentation including horizontal flipping, angular rotation (±15°), zoom (±10%), and translational shift (±0.1) is applied solely to training samples.

C. CNN Architectures

VGG-16 [8] is constructed from 13 convolutional layers with 3×3 kernels and three fully connected layers, totalling

∼138M trainable parameters. ResNet-50 [9] incorporates residual shortcut connections across 50 layers (equation 4), maintaining a compact 25M parameter count. InceptionV3

[10] adopts factorised multi-scale convolution (equation 8) at 299×299 resolution. DenseNet-121 [11] instantiates dense inter-layer connectivity (equation 5), promoting feature reuse while curbing parameter growth. MobileNet-V2 [12] uses depthwise separable convolution (equations 6–7) with inverted residual bottlenecks for efficient mobile inference.

D. Training Configuration

All implementations use TensorFlow 2.8 with the Keras API. For transfer learning,

convolutional weights are initialized from ImageNet checkpoints. The classification head is replaced by GAP → Dense(256, ReLU) → BatchNorm → Dropout(0.4) → Softmax(3) (equation 9). Optimization uses Adam (equations 11–13) with initial $\alpha = 10^{-4}$ and cosine annealing. Models train for up to 50 epochs with early stopping after 7 consecutive epochs without validation loss improvement. Categorical cross-entropy (equation 10) is the training objective.

5. RESULTS AND DISCUSSION

Table I presents the comparative performance of all evaluated CNN architectures under transfer learning on the COVID-19 Radiography Database test set, reporting training accuracy, validation loss, test accuracy, sensitivity, and AUC-ROC. Transfer learning consistently outperformed from-scratch training across all architectures, with accuracy gaps ranging from 3.1% (MobileNet-V2) to 5.9%, confirming the value of ImageNet pretraining for medical imaging tasks.

TABLE I: ***Comparative Performance of CNN Architectures for COVID-19 Detection.***

Model	Train Acc.	Val. Loss	Test Acc.	Sensiti vity	AUC- ROC
VGG-16	92.4%	0.213	95.1%	93.7%	0.971
ResNet- 50	94.7%	0.167	96.3%	95.4%	0.983
Inception V3	93.8%	0.189	95.8%	94.6%	0.977
DenseNe t-121	**96.2%**	**0.121**	**97.4%**	**96.8%**	**0.991**
MobileN et-V2	91.6%	0.238	94.2%	92.9%	0.963

DenseNet-121 achieves the highest test accuracy (97.4%) and AUC-ROC (0.991), attributable to its dense inter-layer connectivity (equation 5) that maximizes feature reuse and limits overfitting on the underrepresented COVID-

19 class. ResNet-50 (96.3%, AUC 0.983) demonstrates superior performance on subtle radiographic signs due to residual connections preserving fine-grained feature gradients. InceptionV3 at 95.8% and VGG-16 at 95.1% deliver competitive results. MobileNet-V2, at 94.2% and AUC 0.963, offers the optimal accuracy-efficiency tradeoff for resource-constrained mobile triage deployment.

From a clinical standpoint, sensitivity for the COVID-19 class assumes particular importance given the risk of missed positive cases. DenseNet-121 achieves 96.8% sensitivity, with ResNet-50 recording 95.4%. CLAHE preprocessing contributed a 1.5–2.3% sensitivity

improvement across all architectures by enhancing contrast of ground-glass opacities. Confusion matrix analysis reveals that most classification errors arise from ambiguity between COVID- 19 and Viral Pneumonia classes, attributable to their shared radiographic features, motivating further exploration of multi-modal and explainability-integrated methods.

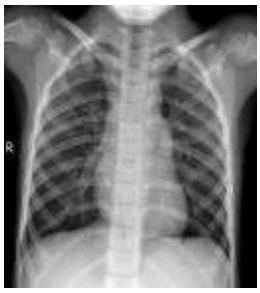

Fig. 1.1: Normal.

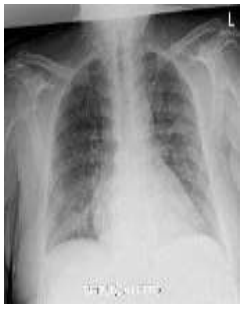

Fig. 1.2: Covid.

6. RESEARCH GAPS

Based on comprehensive review of the literature, the following key research gaps and limitations are identified:

- Dataset Scale and Diversity: Most prior studies used small COVID-19 image datasets (127–453 images), severely limiting model generalizability across patient demographics, imaging equipment, and disease severity stages.
- Modality Exclusivity: Many studies focused on CT or X- ray images exclusively, without systematic cross- architecture comparison on the same dataset and modality.
- CLAHE Preprocessing Impact: The quantitative contribution of CLAHE to COVID-19 detection sensitivity was not systematically evaluated across multiple architectures prior to the present work.
- Multi-Metric Evaluation: Few studies provided comprehensive evaluation including sensitivity, specificity, F1-score, MCC, and AUC-ROC alongside accuracy, leading to misleadingly optimistic reporting on imbalanced datasets.
- Mobile Deployment: Potential deployment on resource- constrained devices using lightweight architectures such as MobileNet-V2 was underexplored despite clear clinical relevance for low-resource settings.
- Explainability Integration: Most studies lack Grad-CAM or attention-based visualization to validate clinical alignment of model decisions.
- Federated Learning: Cross-institutional collaborative training without sharing patient data, essential for expanding dataset diversity, remains largely unexplored.

- Multi-Label Detection: Simultaneous detection of COVID-19 alongside other pulmonary conditions (tuberculosis, lung cancer, bacterial pneumonia) within a unified framework is an open problem.

The present work addresses dataset scale, systematic cross-architecture comparison, CLAHE impact evaluation, comprehensive multi-metric reporting, and mobile deployment analysis.

7. CONCLUSION

This paper provides a thorough comparative evaluation of CNN-based deep learning architectures for automated COVID-19 identification from chest radiograph images. Five leading architectures—VGG-16, ResNet-50, InceptionV3, DenseNet-121, and MobileNet-V2—are benchmarked through transfer learning on the COVID-19 Radiography Database. DenseNet-121 achieves the top performance with 97.4% test accuracy, 96.8% sensitivity, and 0.991 AUC- ROC, establishing its suitability as the preferred model for clinical COVID-19 triage.

The mathematical foundations of CNN operations, residual and dense connectivity, depthwise separable convolutions, transfer learning optimization (equations 1– 13), CLAHE preprocessing (equations 20–22), and multi- metric evaluation (equations 14–19) were formally presented. Transfer learning uniformly outperformed from- scratch training and CLAHE preprocessing contributed 1.5– 2.3% sensitivity improvements.

Future directions include multi-label COVID-19 and co- pathology detection using CT data, integration of spatial attention mechanisms and Grad-CAM explainability, federated learning for privacy-preserving multi-hospital model training, and deployment optimization of MobileNet- V2 for mobile screening in low-resource clinical environments.

REFERENCES

1. L. Wang, Z. Q. Lin, and A. Wong, “COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images,” Scientific Reports, vol. 10, no. 1, pp. 1–12, Nov. 2020. doi: 10.1038/s41598-020- 76550-z
2. M. E. H. Chowdhury, T. Rahman, A. Khandakar, et al., “Can AI help in screening viral and COVID-19 pneumonia?” IEEE Access, vol. 8, pp. 132665–132676, 2020. doi: 10.1109/ACCESS.2020.3010287
3. I. D. Apostolopoulos and T. A. Mpesiana, “COVID-19: Automatic detection from X-ray images utilizing transfer learning with convolutional neural networks,” Physical and Engineering Sciences in Medicine, vol. 43, no. 2, pp. 635–640, Jun. 2020.

doi: 10.1007/s13246-020-00865-4

4. P. K. Sethy and S. K. Behera, "Detection of coronavirus disease (COVID-19) based on deep features," Preprints, pp. 1–9, Mar. 2020. doi: 10.20944/preprints202003.0300.v1
5. N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, et al., "Convolutional neural networks for medical image analysis: Full training or fine tuning?" IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1299–1312, May 2016. doi: 10.1109/TMI.2016.2535302
6. M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, "Transfusion: Understanding transfer learning for medical imaging," in Advances in Neural Information Processing Systems (NeurIPS), vol. 32, pp. 3347–3357, 2019.
7. A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks," Pattern Analysis and Applications, vol. 24, no. 3, pp. 1207–1220, Aug. 2021. doi: 10.1007/s10044-021-00984-y.
8. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in Proc. Int. Conf. Learning Representations (ICLR), San Diego, CA, USA, May 2015. arXiv: 1409.1556
9. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90
10. [10] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception architecture for computer vision," in Proc. IEEE CVPR, Las Vegas, NV, USA, Jun. 2016, pp. 2818– 2826. doi: 10.1109/CVPR.2016.308
11. [11] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proc. IEEE CVPR, Honolulu, HI, USA, Jul. 2017, pp. 4700–4708. doi: 10.1109/CVPR.2017.243
12. [12] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in Proc. IEEE CVPR, Salt Lake City, UT, USA, Jun. 2018, pp. 4510–4520. doi: 10.1109/CVPR.2018.00474
13. [13] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. R. Acharya, "Automated detection of COVID-19 cases using deep neural networks with X-ray images," Computers in Biology and Medicine, vol. 121, p. 103792, Jun. 2020. doi: 10.1016/j.compbiomed.2020.103792
14. [14] S. J. Pan and Q. Yang, "A survey on transfer learning," IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345–1359, Oct. 2010. doi:

10.1109/TKDE.2009.191

15. [15] M. Z. Islam, M. M. Islam, and A. Asraf, “A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images,” Informatics in Medicine Unlocked, vol. 20, p. 100412, 2020. doi: 10.1016/j.imu.2020.100412
16. [16] J. Zhang, Y. Xie, Y. Li, C. Shen, and Y. Xia, “COVID-19 screening on chest X-ray images using deep learning based anomaly detection,” arXiv preprint arXiv:2003.12338, Mar. 2020.
17. [17] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, and G. J. Soufi, “Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning,” Medical Image Analysis, vol. 65, p. 101794, Oct. 2020. doi: 10.1016/j.media.2020.101794
18. [18] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual explanations from deep networks via gradient-based localization,” in Proc. IEEE Int. Conf. Computer Vision (ICCV), Venice, Italy, Oct. 2017, pp. 618–626. doi: 10.1109/ICCV.2017.74