

Improving Techniques for Convolutional Neural Networks Performance

Dina Darwish*

ABSTRACT

Convolutional Neural Networks (CNNs) have been extensively used in several application domains. Researchers have been exploring methods to enhance the accuracy of applications in accuracy-critical domains by either increasing the depth or width of the network. The presence of structures results in a significant increase in both computational and storage costs, hence causing a delay in response time. Convolutional Neural Networks have significantly contributed to the rapid development of several applications, including image classification, object detection, and semantic segmentation. However, in some applications that need zero tolerance for mistakes, such as automated systems, there are still certain issues that need to be addressed to achieve better performance. Then, despite the progress made so far, there are still limitations and challenges that must be overcome. Simultaneously, there is a need for reduced reaction time. Convolutional Neural Networks (CNNs) are now faced with significant obstacles of a formidable nature. This paper investigates different methods that can be used to improve convolutional neural network performance.

Keywords: Accuracy, Convolutional Neural Networks, Improving Techniques, Performance.

Submitted: November 29, 2023

Published: January 22, 2024

 10.24018/ejece.2024.8.1.596

Ahram Canadian University, Egypt.

*Corresponding Author:
e-mail: dina.g.darwish@gmail.com

1. INTRODUCTION

A Deep Learning method called a Convolutional Neural Network (ConvNet/CNN) can take an input picture, apply weights and biases to various objects, and distinguish them. ConvNets need less pre-processing than other classification techniques. With training, ConvNets can learn characteristics that are hand-engineered in basic approaches. A ConvNet's design is similar to the brain's neuron connection pattern and was inspired by the Visual Cortex. Neural responses are limited to the Receptive Field of the visual field. These fields overlap to span the whole visual field.

Since 1989, Convolutional Neural Networks (CNNs) have excelled in semantic segmentation, picture classification, and object identification. Classic models have shown that CNN depth affects performance [1]. Many visual recognition tasks have also benefited from deep networks [2]. As hardware has improved, CNN performance has grown tremendously because of its computing capacity. Unfortunately, most CNNs are not as accurate as human vision. Recent attempts to increase CNN performance include regularizing parameters [3], using improved loss

functions [4], altering pooling procedures [5], and developing more meaningful net task topologies [6]. A deeper network often performs better than a shallower network, and increasing the depth may simply provide a superior model. However, deep CNNs have several drawbacks. First, disappearing or bursting gradients in a deeper network might cause findings to diverge [7]. In addition, deeper networks have more parameters. In the late phases of networks, the number of parameters astronomically grows as convolution channels expand, increasing processing expenses. Due to overfitting, additional stacked layers slow network speed. Specifically, the model has reduced training error but larger testing error. As network depth rises, accuracy saturates and declines quickly [2].

Several approaches proved to improve performance or reduce computational resources and reaction time. On one side, ResNet-150 or even ResNet-1000 has been suggested to enhance extremely little performance margin with enormous computational cost. With a pre-defined performance loss relative to best-effort networks, many strategies have been suggested to reduce compute and storage to fit hardware implementation limits. Quantization, pruning, and lightweight network architecture [8], [9] are

examples. Knowledge Distillation (KD) [10] was a model compression method or tactic.

COVID-19, a novel coronavirus, was verified in humans in late 2019 and causes respiratory issues, heart infections, and mortality. Medical pictures may help prevent COVID-19 spread and treat patients to minimize mortality [11]. Chest X-ray radiography and CT are used in clinical practice to identify, evaluate, and monitor COVID-19. Although CT has greater detection sensitivity, chest X-ray radiography is more often employed in clinical practice because of its inexpensive cost, low radiation dosage, ease of use, and broad accessibility in general or community hospitals [12]. However, several viruses and bacteria may cause pneumonia. Thus, general radiologists at community hospitals may struggle to interpret a large number of chest X-ray pictures to identify mild COVID-19-infected pneumonia and distinguish it from other community-acquired pneumonia. Because COVID-19-infected pneumonia is comparable to other viruses and bacteria. Thus, radiologists encounter this clinical problem during the pandemic [13]. Computer-aided detection or diagnosis (CAD) schemes based on medical image processing and machine learning have garnered research interest to automatically analyze disease characteristics and provide radiologists with decision-making tools for more accurate and efficient COVID-19 pneumonia detection and diagnosis.

Detecting COVID-19 pneumonia may include many steps, including preprocessing of images, segmentation of regions of interest (ROIs) associated to diseases, calculation and selection of relevant image features, and the development of machine learning models that use multiple-feature fusion techniques to discover and classify instances. A study [14] produced 961 image features by analyzing segmented regions of interest (ROIs) in chest X-ray images. A feature selection methodology was used to construct a K-nearest neighbors (KNN) classification model, which effectively categorized cases as either COVID-19 or non-COVID-19 with a high accuracy rate of 96.1%. Recent studies have demonstrated that the development of computer-aided diagnosis (CAD) schemes utilizing deep learning algorithms, without the need for segmenting suspicious regions of interest (ROIs) or calculating hand-crafted image features, is a more efficient and reliable approach compared to traditional machine learning methods. This is primarily due to the challenges associated with accurately identifying and segmenting subtle pneumonia-related disease patterns on chest X-ray images. Therefore, a multitude of deep-learning models have been documented in the literature to detect and categorize COVID-19 [15]. CNN models have been used for CT images [16]; however, a significant body of research has mostly utilized these models to identify and categorize COVID-19 in chest X-rays. The provided models consist of Resnet50 [17], MobileNetV2 [18], CoroNet [19], Xception + ResNet50V2 [20], and several other distinct CNN models. The subsequent rest sections of this paper are structured as follows; section 2 presents the methodology of work of convolutional neural networks, and Section 3 compares several techniques aimed at enhancing the performance of convolutional networks. Lastly, the paper concludes with a conclusion and references.

2. CONVOLUTIONAL NEURAL NETWORKS METHODOLOGY

CNN architecture resembles a brain connection. CNNs contain billions of neurons organized in a precise fashion, like the brain. CNN neurons are organized like the brain's frontal lobe, which processes visual inputs. This configuration covers the complete visual field, unlike standard neural networks, which must be given reduced-resolution pictures. CNNs perform better with image and voice inputs than earlier networks.

A deep-learning CNN has three layers: convolutional, pooling, and Fully Connected (FC). The first layer is convolutional and the final is FC. CNN complexity grows from convolutional to FC layers. This rising complexity helps the CNN to detect greater chunks and more complicated aspects of a picture until it finds the full item.

2.1. Convolutional Layer

CNNs' primary convolutional layer does most calculations. A second convolutional layer may follow the first. A kernel or filter within this layer checks for features in the image's receptive fields during convolution. If we have an input with the dimensions $W \times W \times D$, and the number of kernels is **Dout**, and the spatial size is F , and the stride is S , and the amount of padding is P , then the following formula may be used to compute the size of the volume of the output:

$$w_{out} = \frac{w - F + 2P}{S} + 1 \quad (1)$$

Formula (1) for convolution layer. This will yield an output volume of size $W_{out} \times W_{out} \times D_{out}$. The kernel iterates through the whole picture. The input pixels and filter dot product are computed after each cycle. The dots provide a feature map or convolved feature. This layer converts the picture into numerical data so the CNN can analyze and extract patterns.

2.2. Pooling Payer

Like the convolutional layer, the pooling layer sweeps a kernel or filter over the input picture. Unlike the convolutional layer, the pooling layer minimizes input parameters and loses information. This layer simplifies and boosts CNN efficiency. The formula for determining the size of the output volume, given an activation map of size $W \times W \times D$, a pooling kernel of spatial size F , and stride S , is as follows:

$$w_{out} = \frac{w - F}{S} + 1 \quad (2)$$

Formula (2) for pooling layer. This yields an output volume of size $W_{out} \times W_{out} \times D$. Pooling offers a degree of translation invariance in all instances, hence ensuring the recognition of an item irrespective of its location inside the frame.

2.3. Fully Connected Layer

CNN image categorization uses characteristics from previous layers in the FC layer. Fully linked implies all inputs or nodes from one layer are connected to every activation unit or node of the following layer.

Non-linearity layers refer to the components inside a neural network that introduce non-linear transformations to the input data. These layers are crucial. Given that convolution is a linear operation and pictures exhibit non-linear characteristics, it is common practice to include non-linearity layers immediately after the convolutional layer. This is done to inject non-linearity into the activation map. There are several categories of non-linear operations, with some of the well-recognized ones being:

- **The sigmoid function** is a mathematical function often used in several fields, such as mathematics, statistics, and machine learning. The sigmoid non-linearity may be represented mathematically as $\sigma(\kappa) = 1/(1 + e^{-\kappa})$. The function maps a real-valued number onto the interval $[0, 1]$ by a process of compression. Nevertheless, one notable drawback of the sigmoid function is its tendency to produce gradients close to zero when the activation is located at either extreme end. In the context of backpropagation, when the local gradient diminishes significantly, it may essentially render the gradient inactive. Additionally, in the scenario when the input data received by the neuron is consistently positive, the sigmoid function will provide either entirely positive or entirely negative outputs. Consequently, this will lead to a fluctuating pattern of gradient updates for the weight parameter, resembling a zig-zag motion.
- **The hyperbolic tangent function**, often denoted as \tanh , is a mathematical function that maps a real-valued integer to the interval $[-1, 1]$. Similar to the sigmoid function, the activity of the mentioned neuron reaches saturation. However, unlike sigmoid neurons, the output of this neuron is centered around zero.
- **The Rectified Linear Unit (ReLU)** is a commonly used activation function in deep learning models. The Rectified Linear Unit (ReLU) has gained significant popularity in recent years. The function $i(\kappa)$ is computed as the maximum value between zero and κ . Put otherwise, the activation may be seen as a threshold set at zero.

When comparing the activation functions sigmoid and \tanh to the rectified linear unit (ReLU), it can be seen that ReLU offers more reliability and significantly enhances the rate of convergence, achieving a six-fold acceleration. Regrettably, one drawback of the Rectified Linear Unit (ReLU) activation function is its susceptibility to fragility during the training process. When a substantial gradient passes through a neuron, it might result in an update that prevents additional updates from occurring. Nevertheless, it is possible to address this issue by establishing an appropriate learning rate. All CNN layers are not completely linked to avoid a dense network. It would increase losses, lower output quality, and be computationally costly.

Each layer of a CNN learns to recognize picture characteristics. Each picture is filtered or kernelled to provide a better, more detailed output after each layer. Lower-layer filters might be basic features. The complexity of filters

increases at each layer to identify distinct properties of the input objects. The subsequent layer's input is derived from the convolved image output, which represents the partially recognized image at each layer. The last fully connected (FC) layer of the convolutional neural network (CNN) is responsible for recognizing and classifying the image or object. The process of convolution is used to filter the input image. Every filter functions by activating certain visual elements inside an image and then transmitting its processed output to the subsequent layer. Each layer inside the network acquires the ability to identify certain features and then iterates this process through several levels, ranging from dozens to hundreds or even thousands. Ultimately, comprehensive item recognition is achieved by the use of many layers of visual data inside the CNN.

The primary concern with conventional neural networks is their lack of scalability. A typical neural network may be suitable for processing tiny images that have a limited number of color channels. The computational capacity and resource requirements necessary for the processing of larger and more intricate images need the use of a larger and more expensive neural network. Throughout training, the neural network exhibits overfitting, when it acquires an excessive amount of information from the training data. Additionally, the model has the potential to acquire data noise, which might have an impact on its performance when tested with new data. The neural network (NN) is unable to identify and analyze data attributes or patterns, as well as the item in question.

In contrast, Convolutional Neural Networks (CNNs) exhibit parameter sharing. The nodes inside each Convolutional Neural Network (CNN) layer are interconnected. Convolutional Neural Networks (CNNs) use a mechanism known as parameter sharing, whereby the weights of the network remain constant as the filters of the layers traverse the input image. The computing requirements of CNN are lower compared to NN. The use of Convolutional Neural Networks (CNNs) offers many advantages in the field of deep learning. Deep learning is a subfield within the broader domain of machine learning, characterized by the use of neural networks consisting of a minimum of three layers. The accuracy of multiple-layer networks surpasses that of single-layer networks. Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) are used in the field of deep learning, with their selection contingent upon the specific application at hand.

Convolutional Neural Networks (CNNs) are highly suitable for tasks such as picture recognition, classification, and computer vision (CV) applications due to their ability to provide precise outcomes, particularly when enough data is available. As the object data traverses through the convolutional neural network's several layers, it progressively acquires knowledge about its characteristics via iterative learning. The process of direct (and deep) learning eliminates the need for manual feature extraction. Convolutional Neural Networks (CNNs) can acquire knowledge about novel recognition tasks and enhance their performance by using pre-existing network structures. These characteristics facilitate the utilization of Convolutional Neural Networks (CNNs) in practical

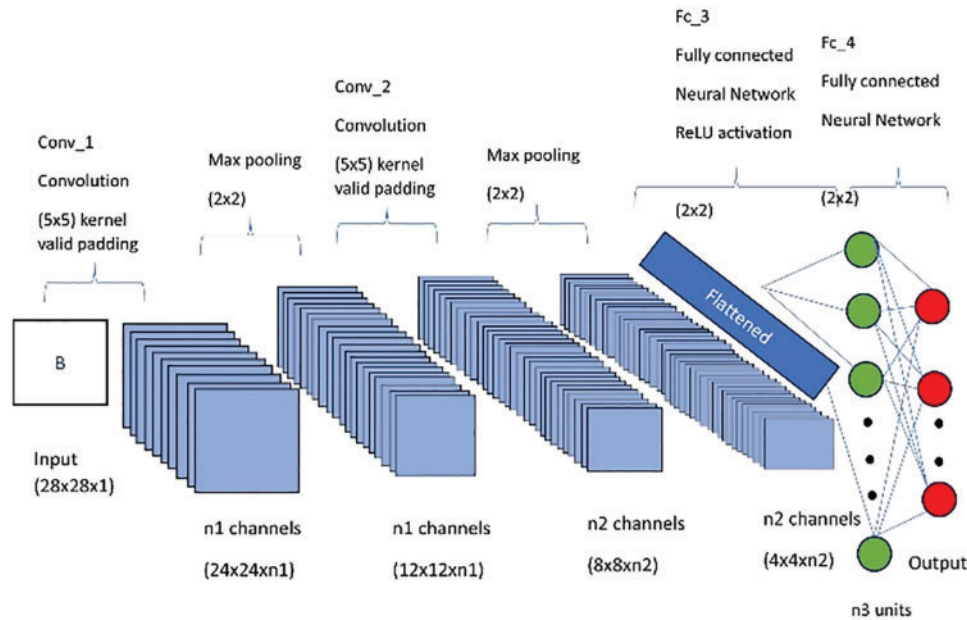


Fig. 1. Convolutional neural network.

scenarios without augmenting computational intricacy or cost. Convolutional Neural Networks (CNNs) use parameter sharing, which enhances their computing efficiency compared to traditional neural networks (NNs). The installation and use of these models on smartphones and other devices are easily comprehensible. Fig. 1 shows a convolutional neural network.

The below section comprises a compilation of apps that now use convolutional neural networks.

1. Object identification has been significantly enhanced with the advent of Convolutional Neural Networks (CNN), which have facilitated the development of sophisticated models such as R-CNN, Fast R-CNN, and Faster R-CNN. These models function as the principal framework for several object recognition models used in autonomous vehicles, facial identification, and various other applications.

2. A team of scholars from Hong Kong developed a Convolutional Neural Network (CNN)-based Deep Parsing Network to enhance an image segmentation model by integrating comprehensive information. Furthermore, a group of academics affiliated with the University of California, Berkeley, have made advancements in the field of semantic segmentation by developing fully convolutional networks that surpass the existing state of the art.

3. The process of generating captions for photos and videos involves the use of Convolutional Neural Networks (CNNs) and recurrent neural networks. This technology demonstrates use across a diverse range of applications, including activity recognition and the provision of audio descriptions for visually impaired individuals in the context of films and images. YouTube has used this technique extensively to establish coherence among the vast quantity of videos uploaded to the platform regularly.

3. COMPARING DIFFERENT TECHNIQUES FOR IMPROVING CONVOLUTIONAL NEURAL NETWORKS PERFORMANCE

There exist several techniques for improving Convolutional Neural Networks performance. This section highlights the major techniques that can be implemented to achieve this goal. These techniques can be listed as follows:

3.1. Adjust Parameters

The epochs, learning rate, etc., may be adjusted to optimize CNN model performance. Epoch count affects performance. Performance improves throughout several epochs. Choosing epochs and learning rates requires experimenting. After a few epochs, training loss does not decrease and accuracy does not increase. Thus, we may choose epochs. CNN model may employ a dropout layer. There is a need to select an optimizer for model compilation based on application. SGD, rmsprop, etc., may be used for optimization. Model tuning requires many optimizers. All of this impacts CNN's performance.

3.2. Image Data Augmentation

CNN needs plenty of training data to automatically learn features from the data. What if we have fewer training data? Image Augmentation solves it. Image augmentation factors like zoom, shear, rotation, preprocessing function, and others enhance data sample count. Using these parameters generates pictures with these properties during Deep Learning model training. Images created by image augmentation often enhance data samples by 3x to 4x. Another benefit of data augmentation is that because CNN is not rotation invariant, pictures may be added by considering rotation. It will boost system precision.

3.3. Deeper Network Topology

A broad neural network may be trained with any input. These networks excel in memory but struggle at generalization. Using a broad, shallow network has certain drawbacks. Wide neural networks can take any input value,

but in reality, we won't have all values for training. Deep networks capture nature's inherent "hierarchy". For example, a convent captures low-level characteristics in the first layer, somewhat better in the following layer, and object pieces and basic structures in subsequent levels. Multiple layers may learn characteristics at different abstraction levels. This is why a deep network may be better than a broad but shallow one. Why not a deep, broad network? The network should be minimal to get decent outcomes. Wider networks take longer to train. Deep networks are computationally costly to train. There is a need to make them broad and deep enough to function, but no wider or deeper.

3.4. Handle Over-and Under-Fitting

This section discusses overfitting and underfitting. A model is a system that converts input to output. We can create an image classification model that predicts class labels for test input images. The data are split into training and testing sets to create a model. On the training set, CNN is used to train a model. The test data output can be forecasted using a trained model.

An overfitted model models training data too effectively. In overfitting, the model has good accuracy on training data but poor accuracy on test data. This means overfitting models are effective at memorizing but not generalization. The model doesn't generalize effectively from training to unseen data. Underfitting is a model that performs poorly on train and test data. The risk is great. Training data does not fit the model properly.

An overfitted model has low bias and large variance. Underfitted models are biased and low-variance. There is a need to trade off bias and variance when building models to find the optimum balance. Bias refers to the presence of mistakes in relation to the training set. Variance refers to the extent to which a model exhibits differences in its predictions when trained on different datasets. The significance of variation lies in the model's inability to accurately predict outcomes when tested against new data.

3.5. Data Normalization

The process of data normalization refers to the systematic organization and structuring of data in a database to eliminate redundancy and improve efficiency. The image tensors were normalized using the process of removing the mean and dividing by the standard deviation of pixels within each channel. The process of normalizing the data ensures that the pixel values in each channel do not have a disproportionate impact on the losses and gradients.

3.6. Batch Normalization

Batch normalization is a technique used in machine learning to improve the performance and stability of neural networks. Following each convolutional layer, a batch normalization layer was included to normalize the outputs of the preceding layer. This process has a resemblance to data normalization, although with a focus on the outputs of a layer. Notably, the mean and standard deviation are treated as parameters that are learned during the training phase.

3.7. Data Augmentation

The process of data augmentation involves using several techniques to expand the size and diversity of a given dataset, without introducing new information. Random transformations were implemented throughout the process of importing photos from the training dataset. In this particular approach, the images will be expanded by adding a border of 4 pixels around each picture. Subsequently, a random section of the image, measuring 32×32 pixels, will be extracted. Additionally, there is a 50% chance that the image will be horizontally flipped.

3.8. Learning Rate Scheduling

In lieu of using a constant learning rate, a learning rate scheduler can be used to dynamically adjust the learning rate after each batch of training. There are several ways for adjusting the learning rate during the training process, for example, the "One Cycle Learning Rate Policy".

3.9. Weigh Decay

Weight decay was included into the optimizer as a means of regularization, serving to mitigate the issue of excessively high weights. This was achieved by introducing an extra term into the loss function.

3.10. Gradient Clipping

Gradient clipping is used as a means to restrict the magnitude of gradients within a narrow range. This approach serves to mitigate the potential adverse effects on model parameters that may arise from too high gradient values during the training process.

3.11. Adams Optimizer

The Adam optimizer is used in lieu of stochastic gradient descent (SGD) to expedite the training process by including momentum and adaptive learning rates. There exists a wide array of alternative optimizers that one might choose to engage in a formal evaluation and experimentation process.

A lot of these techniques can be combined to improve the performance of the Convolutional Neural Networks (CCN) and increase accuracy. Besides, a lot of research papers discussed the issue of improving the performance of the Convolutional Neural Networks (CCN).

For example, In the paper [21], in light of the research conducted by Zhang *et al.* [22], the authors have undertaken a series of tests to demonstrate the ability of the self-distillation framework suggested in this study to converge towards a flat minimum. Two 18-layer Residual Networks (ResNets) were trained on the CIFAR100 dataset, with one using self-distillation and the other without. Next, Gaussian noise is introduced to the parameters of both models. The resulting entropy loss and projected accuracy on the training set are then computed and visualized. The training set accuracy of the model trained with self-distillation remains consistently high as the noise level, represented by the standard deviation of the Gaussian noises, increases. Conversely, the training accuracy of the model without self-distillation experiences a significant decline.

The CIFAR100 dataset, as described in reference [23], comprises of RGB pictures that are small in size, measuring 32×32 pixels. The dataset consists of 100 classes, with a training set of 50,000 photographs and a testing set including 10,000 images. The kernel sizes and strides of neural networks are modified to accommodate the dimensions of small-scale pictures. The ImageNet2012 classification dataset [24] comprises 1,000 classes based on the WordNet taxonomy. Every class is represented by a multitude of pictures. The photos are resized to dimensions of 256×256 pixels, using RGB color representation. It should be noted that the claimed accuracy of ImageNet is calculated based on the validation set.

Self-distillation significantly enhances the performance of Convolutional Neural Networks, yielding substantial improvements without compromising reaction time. On average, an accuracy boost of 2.65% is achieved, with a minimum increase of 0.61% in ResNet and a maximum increase of 4.07% in VGG19. Self-distillation enables the generation of a single neural network executable with varying depths, hence allowing for adaptive trade-offs between accuracy and efficiency on edge devices with restricted resources. Experiments were done to validate the generalization of convolutional neural networks over five different architectures and two distinct datasets. Deeper classifiers in self-distillation provide a greater extraction of discriminating characteristics. The presence of many classifiers in self-distillation allows for the computation and analysis of the characteristics of each classifier, therefore demonstrating their respective discriminating principles.

In another study [25], a proposed methodology is presented to enhance performance via the integration of low-level information derived from several blocks. The authors used five distinct convolutional operations, namely 3×3 , 5×5 , 7×7 , $5 \times 3 \cup 3 \times 5$, and $7 \times 3 \cup 3 \times 7$, to provide five low-level characteristics. Additionally, they proposed two fusion techniques: low-level feature fusion (L-Fusion) and high-level feature fusion (H-Fusion). The experimental findings indicate that the use of L-Fusion yields more efficacy in enhancing the performance of Convolutional Neural Networks (CNNs). Additionally, it is seen that the 5×5 convolution technique is better suited for the purpose of multiscale feature fusion. The conclusion is summarized as a strategic approach that integrates multiscale characteristics in the previous stage of Convolutional Neural Networks (CNNs). In addition, the authors provided a novel architectural approach for the interpretation of Convolutional Neural Networks (CNNs) inputs, using two self-regulating blocks that are guided by a specific strategy. In this study, they made modifications to five pre-existing networks, namely Dense-Net-BC with a depth of 40, ALL-CNN-C with a depth of 9, Darknet 19 with a depth of 19, Resnet 18 with a depth of 18, and Resnet 50 with a depth of 50.

The fusion procedures described in the first stage were assessed by the authors using two well-recognized benchmark datasets, namely CIFAR10 and CIFAR100 [26]. Both datasets consist of 50,000 training photos and 10,000 test images. However, the first dataset is composed of 10 categories, while the second dataset has 100 categories. During the training stage, all 50,000 training photos were

used for training purposes without including a validation set. Subsequently, during the testing stage, the 10,000 test images were employed for evaluation. The data was normalized by using the channel means and standard deviations, as described in the paper [27]. During the training process, the researchers used a data augmentation technique that included random cropping, random horizontal flips, and normalization [28]. This particular technique has been extensively utilized for both datasets [29]–[31], resulting in the creation of two augmented datasets known as CIFAR10+ and CIFAR100+, respectively. In the process of conducting tests, the data was normalized only by the use of channel means and standard deviations [28].

This involves the partitioning of a Convolutional Neural Network (CNN) into distinct blocks based on feature size, to acquire low-level and high-level features for feature fusion. Two fusion methodologies, namely L-Fusion and H-Fusion, were devised in order to evaluate the impact of feature fusion at various phases. To assess the efficacy of multiscale feature fusion, the authors have chosen five low-level characteristics that exhibit varying scales. The technique of L-Fusion has been demonstrated to enhance performance by combining low-level features retrieved by a convolutional neural network (CNN) with various scales derived from an auxiliary block. The auxiliary block may be constructed based on the architecture of the initial block in a Convolutional Neural Network (CNN). The conclusion is validated using a set of five Convolutional Neural Networks (CNNs) that exhibit excellent classification accuracy. The experimental findings demonstrate that the technique attains performance that is currently considered the best in its field. Concurrently, the suggested design does not significantly augment the parameter count of a Convolutional Neural Network (CNN) since the fusion operation occurs in the preceding step.

To construct Darknet-19-Fusion, the authors included two parallel 5×5 convolutional layers. Additionally, they introduced a parallel block that substitutes the 7×7 convolutional layer and all 3×3 convolutional layers with a 5×5 convolutional layer to create Resnet-18-Fusion and Resnet-50-Fusion. It has been observed that the performance of Darknet 19 [32] has seen a minor improvement with the incorporation of two parallel convolutional layers and the use of concatenation and CR fusion procedures. In a similar vein, the inclusion of a parallel block in the previous stage yields notably improved outcomes for Resnet 18-Fusion and Resnet-50-Fusion in comparison to Resnet 18 and Resnet 50 [33]. In the case of Resnet-18-Fusion, there is an enhancement seen in the top-1 accuracy and top-5 accuracy, with an increase of 1.59% and 0.96% respectively. Regarding Resnet-50-Fusion, a reduction of 0.4% in the top-1 error and 0.6% in the top-5 error is seen.

From the results, it is evident that the methodology used yields a significant improvement in the performance of DenseNet-BC and ALL-CNN-C. In the CIFAR10+ dataset, an improvement was seen in the accuracy of DenseNet-BC and ALL-CNN-C by 0.76% and 1.15%, respectively. In the CIFAR100+ dataset, a reduction in error rates was achieved for the DenseNet-BC and ALL-CNN-C models by 2.25% and 4.68% respectively. Despite the significant rise in the number of parameters, the

computing cost remains very low. The ILSVRC 2012 classification dataset was chosen to assess the suitability of the suggested technique for a more extensive dataset. The Darknet 19, Resnet 18, and Resnet 50 models demonstrate strong performance on the ILSVRC 2012 classification dataset, while also exhibiting variations in the number of convolutional layers used. The L-Fusion 1 design serves as the basis for constructing Dark-net 19-Fusion, Resnet-18-Fusion, and Resnet-50-Fusion networks.

In the third study [34], the authors created and evaluated a chest X-ray computer-aided diagnosis (CAD) method to identify COVID-19-infected pneumonia. The CAD system first eliminates a majority of diaphragm regions, followed by the use of a histogram equalization technique and a bilateral low-pass filter to process the image. The composite image is formed by merging the unaltered original photograph with two processed photographs that have undergone filtering techniques, resulting in a fake color image. The classification of chest X-ray images into three categories, namely COVID-19-infected pneumonia, other community-acquired no-COVID-19 pneumonia, and normal (non-pneumonia), is achieved using a convolutional neural network (CNN) model based on transfer learning. A public dataset of 8,474 chest X-ray pictures with 415, 5,179, and 2,880 cases in three classes is utilized to develop and evaluate the CNN model. To train and test the CNN model, the dataset is randomly partitioned into training, validation, and testing subsets with the same frequency of examples in each class. The CNN-based CAD scheme classifies 3 classes with 94.5% accuracy (2404/2544) and a 95% confidence interval of [0.93], [0.96]. CAD had 98.4% sensitivity (124/126) and 98% specificity (2371/2418) in diagnosing COVID-19 patients. CAD has 88.0% classification accuracy without two preprocessing steps (2239/2544). So, adding two image preprocessing stages and producing a pseudo-color picture improves the accuracy of a deep learning CAD approach for chest X-ray images to identify COVID-19-infected pneumonia.

Recent research has demonstrated that the development of computer-aided diagnosis (CAD) systems using deep learning algorithms, which do not require the segmentation of suspicious regions of interest (ROIs) or the computation of handcrafted image features, is more successful and reliable compared to traditional machine learning methods. This is primarily attributed to the challenges associated with accurately identifying and segmenting subtle disease patterns or ROIs related to pneumonia on chest X-ray images. Thus, numerous deep-learning models have been published to identify and classify COVID-19. Some CNN models are used on CT images, but many research uses them on chest X-rays to identify and categorize COVID-19. Resnet50, MobileNetV2, CoroNet, Xception + ResNet50V2 and other additional unique CNN models are included. Different image datasets of 25 to 224 COVID-19 instances out of 50 to 11,302 cases were employed in this research. COVID-19 detection sensitivity was 79.0%–98.6%. Previous research has shown encouraging outcomes, but many questions remain about how to train deep learning models properly. The authors used and compiled a collection of chest X-ray radiography (CXR) pictures from many public

medical archives [35], [36]. The aforementioned repository was established and analyzed by a consortium of esteemed organizations including the Allen Institute for AI, the Chan Zuckerberg Initiative, Georgetown University's Centre for Security and Emerging Technology, Microsoft Research, and the National Library of Medicine-National Institutes of Health, in partnership with the White House Office of Science and Technology Policy. In particular, this research employed 8,474 posteroanterior (PA) chest 2D X-ray scans. 415 pictures show COVID-19, 5179 show community-acquired pneumonia, and 2,880 show normal cases.

The authors created and tested a deep transfer learning CNN model to identify and classify COVID-19-infected pneumonia chest X-rays. This research differs from past studies on this subject and yields some new intriguing findings. The deep learning CNN model has many parameters to train and determine, therefore a big and varied picture dataset is needed to give reliable findings [37]. The authors utilized 8,474 chest X-ray pictures; however, the dataset is uneven in 3 classes, and the number of COVID-19-infected pneumonia patients (415) is modest. Therefore, they used class weight during training, picked a well-trained VGG16 model, and used transfer learning to develop a robust deep-learning model. Over 138,000,000 parameters comprise the original VGG16 model. A 14-million-image ImageNet database trained and determined these parameters. Robustly training so many parameters from the start with 8,474 photos is tough. To prevent overfitting, they retrained or fine-tuned the pre-trained VGG16. Study results show that this transfer learning approach can improve performance with an accuracy of 94.5% (2404/2544) in classifying 3 classes and 98.1% (2495/2544) in classifying cases with and without COVID-19 infection, as well as a Co-hen's kappa score of 0.89. Second, chest X-rays are grayscale, unlike color photos.

To fully leverage the pre-trained VGG16-based CNN model, the authors produced two gray-level pictures. Three gray-level pictures are supplied into three RGB color channels of the CNN model instead of the original chest X-ray image. They used a bilateral low-pass filter to minimize noise and a histogram equalization approach to normalize contrast. Compared to using only the original chest X-ray image and three different images to generate a pseudo-color image as an input image to the CNN model, this study found that using a pseudo-color image approach increases classification accuracy by 3.6% from 91.2%–94.5%, and Cohen's kappa score by 7.2% from 0.83 to 0.89. Using their innovative strategy to fully exploit 3 input channels of the CNN pre-trained model with color pictures improves image classification since these two filtered gray-level images include more information. Third, preprocessing is important because illness patterns in medical imaging are usually unique [38]. Thus, they used an image preprocessing technique to automatically recognize and eliminate the majority of the diaphragm area from chest X-rays. Removing the majority of diaphragm regions increases the CNN model's classification accuracy by 7.4% and Cohen's kappa coefficient by 18.7%. Although deep learning skips the segmentation of suspicious disease regions of interest, this study shows that applying an image

preprocessing and segmentation algorithm to remove irrelevant regions on the image can also improve deep learning model performance and robustness.

In the fourth study [39], Convolutional neural networks (CNNs) have shown encouraging outcomes in the domain of end-to-end voice recognition, while their performance still lags behind that of other cutting-edge approaches. This research examines how the disparity may be addressed and surpassed by the implementation of an innovative architecture known as ContextNet, which combines Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and a transducer. The architecture of ContextNet includes a convolutional encoder that is capable of integrating global context information into convolution layers via the use of squeeze-and-excitation modules. Furthermore, the authors provided a straightforward scaling technique that effectively adjusts the widths of ContextNet, resulting in a favorable balance between computational requirements and accuracy. In this study, the authors provide evidence that ContextNet achieves a word error rate (WER) of 2.1%/4.6% on the internationally recognized LibriSpeech benchmark, both without the use of an external language model (LM) and with the inclusion of an LM. Additionally, they find that ContextNet achieves a WER of 2.9%/7.0% while using just 10 M parameters on the clean/noisy LibriSpeech test sets. This may be compared to the previously reported optimal system performance of 2.0% for the LM approach and 4.6% for the 20 M parameter approach, as well as 3.9% for the LM approach and 11.3% for the 20 M parameter approach. The efficacy of the suggested Context-Net model is further substantiated by its evaluation of a much bigger internal dataset.

In the fifth study [40], the activation function serves as a fundamental element in the convolutional neural network (CNN), since it enables the network to perform nonlinear transformations. Various activation functions are used to enable the original input to contend with distinct linear or nonlinear mapping terms, resulting in diverse nonlinear transformation capabilities. Until recently, the first input of the activation function known as funnel activation (FReLU) was shown to compete with spatial circumstances. As a result, FReLU has the capacity for nonlinear transformation.

However, it also has the capability of pixel-wise modeling. In this analysis, the authors provided a summary of the competitive process inside the activation function, followed by an examination of its implications. This paper presents a fresh design template for an activation function. The competitive activation function (CAF) is an algorithmic mechanism that facilitates the promotion of competition among various elements. The Comprehensive Activation Function (CAF) encompasses a broad range of activation functions that use competitive processes. Based on the findings of the CAF, it is thus proposed that the term “parametric funnel rectified exponential unit (PFREU)” refers to a specific mathematical function or model that incorporates parametric properties, rectification, and an exponential unit. The PFREU framework facilitates the cultivation of competition between linear and nonlinear mappings, and the circumstances related to

space. In this study, the authors performed experiments on four distinct datasets; with varying sizes. These datasets include; Fashion-MNIST, CIFAR; is one of the most widely used color image datasets, and it consists of two subsets: CIFAR-10 and CIFAR-100, and Tiny ImageNet, which is a subset of the ImageNet dataset. For all models, the authors have chosen the Xavier [41] initialization strategy and the classification cross-entropy loss function. Also, they used LeNet-5 [42], the Network in Network (NIN) [43], and the Residual Network (ResNet) [44] to evaluate the performance of different activation functions. Additionally, they analyzed the experimental outcomes of three well-established methodologies. The strategy used in our study was shown to be more effective via the utilization of convolutional neural networks. The FReLU activation function has superior performance in comparison to its predecessors in the realm of nonlinear activation functions. The findings indicate that the inclusion of a spatial condition in FReLU allows for pixel-wise modeling, a capability that is absent in traditional activation functions.

In the sixth study [45], the identification of stenosis in X-ray Coronary Angiography (XCA) pictures using automated means has the potential to facilitate the early diagnosis of coronary artery disease. Stenosis is characterized by the accumulation of plaque inside the arteries, resulting in a reduction of blood flow to the heart and an elevated susceptibility to experiencing a myocardial infarction. Convolutional Neural Networks (CNNs) have shown efficacy in accurately distinguishing diseased, regular, and distinctive tissues across extensive and varied medical picture datasets. However, convolutional neural networks (CNNs) have operational and performance constraints when dealing with tiny and inadequately diverse datasets. The use of transfer learning from large datasets of natural images, such as ImageNet, has emerged as a widely accepted approach to enhance the performance of neural networks in the field of medical imaging.

This study introduces a unique Hierarchical Bezier-based Generative Model (HBGM) as a means to enhance the training process of Convolutional Neural Networks (CNNs) for the purpose of stenosis detection. In this study, the authors propose the generation of fake picture patches as a means to augment the original database, hence expediting the convergence of the network. The synthetic dataset has 10,000 photos, with an equal distribution of 50% depicting stenosis instances and 50% representing non-stenosis cases. In addition, the evaluation of the produced data is quantitatively assessed using a dependable Fréchet Inception Distance (FID). Consequently, the suggested methodology involves pre-training the network using false datasets and later fine-tuning it using the authentic XCA training dataset. The dataset used in this study comprises 250 XCA image patches, with 125 patches representing stenosis instances and the remaining patches representing non-stenosis cases. In addition, the network design included a Convolutional Block Attention Module (CBAM) as a self-attention mechanism to enhance the network's efficiency.

The findings of the study indicate that the pre-trained networks, which used the suggested generative model, exhibited superior performance compared to the results

obtained by training the networks from the beginning. The obtained values for accuracy, precision, sensitivity, and F1-score were 0.8934, 0.9031, 0.8746, 0.8880, and 0.9111 respectively. The artificially manufactured dataset has a mean Fréchet Inception Distance (FID) of 84.0886, indicating a higher level of realism in the visual XCA pictures. Various ResNet topologies have been assessed in the context of stenosis detection, with the incorporation of attention modules within the network being considered. The numerical findings indicate that the use of the HBGM yields superior performance compared to training from the initial state, surpassing even the performance of models pre-trained on ImageNet.

The inclusion of an activation function is a crucial method for incorporating nonlinearity into convolutional neural networks. The activation functions frequently used in various applications mostly utilize several types of negative feedback to handle negative input. Nonetheless, in recent times, some researchers have begun to explore other approaches. One suggested approach for addressing negative input is the use of positive feedback methods, such as the utilization of Con-catenated Rectified techniques. The use of Linear Units (CReLU) and Linearly Scaled Hyperbolic Tangent (LiSHT) has resulted in improved outcomes. The subject of discussion is performance.

In the seventh research [46], in order to go further into this concept, the authors put forward a novel activation function, which they referred to as the difference Exponentially Linear Unit (DELU). The DELU activation function that has been suggested may provide positive and negative feedback for varying negative input values. In this study, the authors examined the experimental outcomes obtained from widely used datasets, including Fashion Mnist, CIFAR10, and ImageNet datasets. The results demonstrated that the DELU activation function has superior performance compared to six other commonly used activation functions, namely Leaky ReLU, ReLU, Exponential Linear Unit (ELU), Scaled Exponential Linear Unit (SELU), Swish, and SERLU.

In the eight study [47], the objective of this research is to provide a comprehensive understanding of the characteristics of convolutional neural networks (CNNs), along with a universal approach to enhance the performance of various CNN designs. In this study, the authors first analyzed the current convolutional neural network (CNN) models and their characteristics. This study aims to investigate an interesting characteristic shown by the filters used in the bottom layers exhibiting a pattern of pairing, namely filters with opposite phases. Motivated by their observations, the authors suggested a unique, efficacious activation method referred to as concatenated Rectified Linear Unit (CReLU), and the reconstruction property of the method under consideration was theoretically analyzed in Convolutional Neural Networks (CNNs). The authors included the CReLU activation function in many state-of-the-art CNN architectures. The activation technique that has been presented maintains both positive and negative phase information, while also ensuring the presence of non-saturated non-linearity. The distinct characteristics of CReLU enable a mathematical representation of convolution layers based on the reconstruction property. This

property serves as a significant measure of the expressiveness and generalizability of the associated features in Convolutional Neural Networks (CNNs).

This study focuses on contemporary Convolutional Neural Networks (CNN) and demonstrated improvements in their recognition capabilities on CIFAR-10/100 and ImageNet datasets using a reduced number of trainable parameters. The findings of this study indicated that a greater understanding of the characteristics of Convolutional Neural Networks (CNNs) has been shown to provide significant improvements in performance via the implementation of a simple tweak.

In the ninth study [48], this research focuses on enhancing the fundamental convolution-al feature transformation process of Convolutional Neural Networks (CNNs) without making adjustments to the model designs. In order to achieve this objective, the authors provided a new self-calibrated convolution technique that effectively enhances the field-of-view of each convolutional layer by means of internal communications, hence enhancing the richness of the output features. In contrast to conventional convolutions that integrate spatial and channel-wise information using small kernels (e.g., 3×3), the self-calibrated convolution employs a unique self-calibration operation to dynamically establish extensive spatial and inter-channel dependencies around each spatial location. Therefore, the incorporation of more information may assist Convolutional Neural Networks (CNNs) in generating more discriminative representations. The self-calibrated convolution architecture presented in this study is characterized by its simplicity and generality. It may be readily used to enhance typical convolutional layers without the need for additional parameters or increased complexity. Experimental studies have shown that the integration of this self-calibrated convolution technique with various backbone models leads to substantial enhancements in multiple visual tasks such as image recognition, object detection, instance segmentation, and key point detection. These improvements are achieved without necessitating any modifications to the underlying network architectures. The authors anticipated that this study will provide valuable insights for future research. This study explored a potentially effective approach to enhance convolutional feature transformation in convolutional neural networks via the development of unique design techniques.

In a tenth study [49], the authors proposed a hybrid methodology aimed at enhancing the precision of Convolutional Neural Networks (CNN) without necessitating model retraining. The suggested architectural modification is the substitution of the SoftMax layer with a K-Nearest Neighbor (kNN) algorithm during the inference process. While this approach is often used in the context of transfer learning when acquiring knowledge, it is customary to use it within the specific field or area in which the neural network was first trained. Prior studies have shown that the neural codes, namely the neuron activations of the deepest hidden layers, may provide advantages from including classifiers such as support vector machines (SVMs) or random forests often used in several academic and practical domains. In the present study, the hybrid approach presented by the research team, the CNN and

kNN architecture is assessed via the use of many picture datasets, network topologies, and label noise levels. The findings demonstrated significant improvements in the accuracy of the inference step when considering the typical Convolutional Neural Network (CNN) architecture in the presence of noisy labels, it becomes very relevant to examine its performance, particularly when dealing with reasonably big datasets like CIFAR100. Additionally, the authors ascertained that the utilization of the '2 norm on neural codes yields statistical advantages for this methodology.

In an eleventh study [50], Texture classification is an issue that may be used in a variety of contexts, such as in the detection of forest species from a distance and the identification of different types of trees. The solutions often are tailored specifically for the dataset utilized, but cannot be generalized in any way. The algorithms for machine learning play a significant part in the way the Texture Classification is currently structured.

These algorithms, on the other hand, suffer from a lack of precision and the rate of categorization. Another kind of sophisticated technology known as the deep learning approach that may overcome these difficulties due to the categorization performance in classical methods is not very good. The combination of a Convolutional Neural Networks Network (CNN) with a Long Short-Term Memory (LSTM) algorithm creates a reliable selection between a robust invariant feature extractor and a classifier that works accurately. It should be possible for the model to automatically decide which attributes are efficient for the texture samples to be created, the feature samples must first be categorized accurately. Fusion performed ensures stability in categorization rates across several data sets and the classification performance of texture will be greatly increased thanks to the suggested model. The results of the experimental investigation indicate that CNN-LSTM achieves superior results as compared to other recent and advanced versions of the algorithms SVM and CNN.

In the twelfth study [51], the authors examined the learning issue associated with one-hidden-layer non-overlapping convolutional neural networks that use the rectified linear unit (ReLU) activation function. The analysis focuses on the perspective of model estimation. The training outputs are hypothesized to be produced by the neural network with the ground-truth parameters that are unknown, together with the inclusion of additive noise. The objective is to estimate the parameters of the model by minimizing a non-convex squared loss function for the training data. Given the assumption that the training set comprises a limited number of samples that have been created from the Gaussian distribution, it can be seen that the accelerated gradient de-scent process may be ensured by using an appropriate starting strategy. The ground-truth parameters are estimated, taking into account the noise level. Despite the non-convex nature of the learning issue, a linear rate is seen. In addition, it has been shown that the convergence rate is quicker than that of the vanilla gradient descent. The process of startup may be accomplished by the current way of initializing tensors. In contrast to the prevailing in theoretical studies, it is assumed that there exists an endless number of samples. The dataset used for

training purposes consists of samples. In this analysis, the neural network under consideration is not deep. This study is an effort to demonstrate the use of accelerated gradient methods in a profound manner. Descent algorithms provide the capability to identify the global optimizer of the non-convex learning issue associated with neural networks. The research conducted focuses on the analysis of the sample complexity associated with gradient-based methods. This paper investigates the various approaches used in the acquisition of knowledge on convolutional neural networks using the non-smooth rectified linear unit (ReLU) activation function. Additionally, this study also offers the strictest upper limit so far on the estimated error with the generated noise, which is known as noise.

In the thirteenth study [52], Despite the allure of deep neural networks as a viable alternative to conventional crafted filters, they nonetheless encounter some special cases that cannot be effectively addressed just by training of convolutional filters, which are a fundamental component of Convolutional Neural Networks (CNNs) used in deep learning. Various atypical elements, such as ambient noise in the real-world setting,

The presence of blur or other forms of quality deterioration significantly impairs the output of a neural network. These unforeseen issues have the potential to give rise to significant complexities, hence leading to crucial difficulties.

Hence, the authors provided a comprehensive examination of the impact of noise on image classification. This inquiry seeks to provide a comprehensive framework for a dual-channel model, while also providing a generalized architecture for its implementation. The objective is to address the issue of deteriorated input photos in terms of their quality. Also, in this research, the authors examined the impact of picture quality on the efficacy of a cutting-edge convolutional network. This study focuses on the challenge of picture categorization and proposes a unique architecture for this purpose. In this study, three prevalent noise types and lossy JPEG compression techniques were selected and analyzed in real-world scenarios.

Many prevalent sources of noise are observable in real-world environments. The types of noise often encountered in image processing include Gaussian noise, salt-and-pepper noise (also known as impulse noise), and speckle noise. The categorization of data in the presence of noise in the surrounding environment is performed. The authors have selected the three most prevalent techniques for input picture de-noising. Many approaches may be used to enhance the robustness of a convolutional neural network model for quality. The rationale for selecting these commonly used denoising techniques is to enhance the edges and outlines of an item while omitting minor particulars. Consequently, it is reasonable to anticipate that the characteristics derived from the denoised and outline-enhanced photos do not contain distortion.

The proposed architecture employs a dual-channel design, using outline-enhanced input mechanisms. An analysis of the minor discrepancies within the suggested architectural design and the techniques used in its training. In order to establish a generalization of the suggested model, the authors conducted an assessment using diverse

sources of varying quality, including an original picture. The hypothesis under consideration pertains to the use of an outline-enhanced picture. In this study, the authors conducted a comparative analysis between the suggested dual-channel model and a basic single model, demonstrating its enhanced performance.

In the fourteenth study [53], the use of Convolutional Neural Network (CNN) has been extensively employed in the domain of medical image processing. Histopathology is a modality used by pathologists to analyze the condition or state of tissues via the examination of pictures.

The picture exhibits an unorganized pattern, resulting in either misidentifying or needing more time for analysis by the medical professional specializing in the study of diseases, known as the pathologist. In addition to this, it is worth noting that the process of training deep learning models often involves, in order to enhance performance, to use of hardware resources of significant capability. Throughout the training program, in order to mitigate these issues, it is essential to speed up the training and optimize the histology dataset. The authors engaged in the process of training CNN using three comparable GPU specifications, namely the GTX-1080, as an alternate approach to expedite the training process. The mean-shift filter is a widely used image processing technique, that uses a low-pass filter. This tool is used for the purpose of enhancing the unstructured pattern on histopathological photos to improve image processing. This technique may be utilized to extract and improve the underlying patterns present in the images, hence improving the overall quality and clarity of the visual representation. By using sophisticated algorithms and methodologies, it is possible to effectively identify and enhance the unstructured pattern. The dataset is a collection of structured and organized data that is used for analysis and research purposes. The performance metrics of all three Graphics Processing Units (GPUs) are shown in the following analysis. The evaluation of the training procedure, which consists of 500 epochs, is conducted by the measurement of speedup. In the meantime, the process of doing model testing is undertaken. There are several possibilities for selecting batch sizes, including 32, 64, 128, and 256. The use of mean-shift has the potential to enhance convergence throughout the training process, with the increase in batch size to 128, resulting in improved convergence.

In a fifteenth study [54], Complex model topologies enable Convolutional Neural Networks (CNNs) to acquire a greater understanding of input characteristics. However, this enhanced learning capability sometimes leads to a decrease in the model's capacity to generalize to unfamiliar data and may result in overfitting issues. Although various regularization techniques. Several techniques have been suggested in the literature to enhance generalization, including data augmentation, batch normalization, and Dropout. Ongoing research aims to further improve the generalization capabilities of machine learning models. The issue of performance remains a prevalent worry throughout the training of robust Convolutional Neural Networks (CNNs). In this paper, the authors provided a novel approach to a dynamically controlled adjustment mechanism, referred to as LossDA, that incorporates a

disturbance variable into the fully connected layer. The trend of this variable exhibits a consistent pattern with the training loss, but the amount of the variable may be predetermined. The system was modified adaptively. The regularization procedure as a whole has the potential to enhance the generalization performance of Convolutional Neural Networks (CNNs) while simultaneously assisting. To mitigate the issue of overfitting, many techniques might be used. To assess the effectiveness of the proposed strategy, this study performs comparison tests on widely used datasets, namely MNIST and FashionMNIST, CIFAR-10, Cats versus Dogs, and miniImagenet, which are often used in academic research. The empirical findings demonstrate that the used approach can enhance the model. This study evaluates the performance of Light CNNs and Transfer CNNs, namely InceptionResNet, VGG19, ResNet50, and InceptionV3. The highest observed enhancement in accuracy for Light CNNs is 4.62%, while the F1 score shows a maximum increase of 3.99%, and the Recall metric demonstrates a maximum improvement of 4.69%. The accuracy increase of Transfer CNNs is 4.17%, the F1 score is 5.64%, and the recall rate is 4.05%. [Table 1](#) summarizes the techniques for improving Convolutional Neural Network (CNN).

From the previous table, it can be concluded that there are techniques to improve the performance of convolutional neural networks, such as:

- Adjust Parameters
- Image Data Augmentation
- Deeper Network Topology
- Handle over- and under-fitting
- Data Normalization
- Batch normalization
- Data Augmentation
- learning rate scheduling
- Weigh decay
- Gradient clipping
- Adams optimizer
- Using Self-distillation
- low-level feature fusion (L-Fusion) and high-level feature fusion (H-Fusion)
- The integration of two additional image preprocessing steps and the generation of a pseudo-color image
- Integrating convolutional neural networks (CNNs), recurrent neural networks (RNNs), and a transducer
- Introducing an activation function, that promotes element competitiveness
- Pre-training the network using bogus datasets and fine-tunes it with the actual training dataset.
- Providing both positive and negative feedback for negative input values
- Explores the odd pairing pattern of filters with opposing phases at the bottom layers
- Introducing a self-calibrated convolution technique that dynamically establishes spatial and inter-channel interdependence
- Replacing the softmax layer with kNN during inference

TABLE I: THE TECHNIQUES FOR IMPROVING CONVOLUTIONAL NEURAL NETWORK

Technique name	Benefits
Adjust parameters	<ul style="list-style-type: none"> • The number of epochs has an impact on performance. • Performance demonstrates improvement throughout several epochs.
Image data augmentation	<ul style="list-style-type: none"> • Image augmentation techniques often result in a significant increase in the data samples, typically by a factor of 3 to 4. • The use of image augmentation has been seen to improve the accuracy of the system.
Deeper network topology	<ul style="list-style-type: none"> • Deeper neural networks are capable of capturing the intrinsic hierarchical structure seen in natural phenomena. • Multiple layers within these networks have the ability to learn and represent properties at various levels of abstraction.
Handle over-and under-fitting	<ul style="list-style-type: none"> • In the phenomenon of overfitting, the model exhibits high accuracy when trained on a specific dataset, then decreased accuracy when applied to new, unseen data. • Underfitting refers to a model's having inadequate performance on both training and testing data. • When constructing models, it is necessary to strike a balance between bias and variance in order to get the optimal outcome.
Data normalization	<ul style="list-style-type: none"> • The term "database normalization" pertains to the methodical arrangement and structuring of data inside a database with the aim of eliminating duplication and enhancing operational efficiency.
Batch normalization	<ul style="list-style-type: none"> • The use of this technique in machine learning serves to enhance the efficacy and robustness of neural networks. • It has a similarity to the process of data normalization.
Data augmentation	<ul style="list-style-type: none"> • The process entails using many methodologies to augment the scale and diversity of a provided dataset, while refraining from adding fresh information.
Learning rate scheduling	<ul style="list-style-type: none"> • A learning rate scheduler is a mechanism that dynamically modifies the learning rate after the completion of each batch of training.
Weigh decay	<ul style="list-style-type: none"> • The use of regularization was included into the optimizer as a method of enhancing the model's performance. • Providing a means to alleviate the problem of extremely elevated weights.
Gradient clipping	<ul style="list-style-type: none"> • The purpose of this technique is to alleviate the possible negative consequences on the model's parameters that might occur due to excessively high gradient values during the training phase.
Adams optimizer	<ul style="list-style-type: none"> • In order to enhance the efficiency of the training process, it is proposed to use the concepts of momentum and adaptive learning rates.
Study 1 [21]	<ul style="list-style-type: none"> • Two residual networks (ResNets) consisting of 18 layers each were trained on the CIFAR100 dataset. One of the ResNets used self-distillation, whereas the other ResNet did not include this technique. Subsequently, the parameters of both models are subjected to the introduction of Gaussian noise. Next, the computation and visualization of the resultant entropy loss and predicted accuracy on the training set are performed. (Experiment) • Self-distillation has been shown to have a considerable impact on the performance of convolutional neural networks, resulting in notable enhancements without any negative effects on response time. (Results)
Study 2 [25]	<ul style="list-style-type: none"> • This paper introduces a suggested technique with the objective of improving performance by integrating low-level information obtained from many blocks. The researchers used five different convolutional processes, namely 3×3, 5×5, 7×7, $5 \times 3 \cup 3 \times 5$, and $7 \times 3 \cup 3 \times 7$, in order to extract five separate low-level features. In addition, two fusion strategies were suggested by the researchers, namely low-level feature fusion (L-Fusion) and high-level feature fusion (H-Fusion). (Experiment) • The experimental results suggest that the use of L-fusion leads to more effectiveness in improving the performance of convolutional neural networks (CNNs). Furthermore, it has been shown that the use of the 5×5 convolution approach is more appropriate for the task of integrating multiscale features. (Results)
Study 3 [34]	<ul style="list-style-type: none"> • The researchers developed and assessed a computer-aided diagnostic (CAD) technique for detecting COVID-19-infected pneumonia in chest X-ray images. The CAD system first eliminates most of the diaphragm regions, followed by using a histogram equalization approach and a bilateral low-pass filter to enhance the image. The composite image is formed by merging the unaltered original photograph with two processed images that have undergone filtering techniques, resulting in a fabricated representation of color. (Experiment) • The integration of two additional image preprocessing steps and the generation of a pseudo color image enhances the efficacy of a deep learning computer-aided diagnosis (CAD) technique for the identification of COVID-19-infected pneumonia in chest X-ray images. (Results)
Study 4 [4]	<ul style="list-style-type: none"> • The use of a novel architectural framework called ContextNet is being employed, which integrates convolutional neural networks (CNNs), recurrent neural networks (RNNs), and a transducer. (Experiment) • The authors have presented a scaling approach that efficiently modifies the widths of ContextNet, achieving a desirable trade-off between computing demands and accuracy. (Results)
Study 5 [5]	<ul style="list-style-type: none"> • This study has summarized and analyzed the competitive process of the activation function and its effects. This research suggests a novel activation function template called "CAF". CAF promotes element competitiveness. According to the CAF, "parametric funnel rectified exponential unit (PFREU)" is a mathematical function or model having parametric characteristics, rectification, and an exponential unit. (Experiment) • Researchers evaluated four datasets of varied sizes. They also examined three well-established methodologies in their experiments. The study's method functioned better using CNNs. (Results)

TABLE I: CONTINUED

Technique name	Benefits
Study 6 [45]	<ul style="list-style-type: none"> • The authors present a novel hierarchical bezier-based generative model (HBGM) to improve CNN training for stenosis identification. In this research, the authors suggest generating bogus image patches to expand the database and speed network convergence. The 10,000-photo synthetic dataset shows 50% stenosis and 50% non-stenosis. The collected data is also statistically evaluated utilizing a reliable fr�chet inception distance. Thus, the recommended strategy pre-trains the network using bogus datasets and fine-tunes it with the actual XCA training dataset. The network also used a convolutional block attention module (CBAM) for self-attention to boost efficiency.(Experiment) • The research found that pre-trained networks using the recommended generative model performed better than those built from scratch. For stenosis identification, ResNet topologies with attention modules have been tested. The numerical results show that the HBGM outperforms training from the starting state and ImageNet-trained models. (Results)
Study 7 [46]	<ul style="list-style-type: none"> • The differential exponentially linear unit (DELU) activation function was presented by the authors. The DELU activation function, if implemented, has the potential to provide both positive and negative feedback for negative input values. (Experiment) • The present study examined empirical findings obtained from the use of fashion Mnist, CIFAR10, and ImageNet datasets. The performance of DELU was superior than that of Leaky ReLU, ReLU, exponential linear unit (ELU), scaled exponential linear unit (SELU), Swish, and SERLU. (Results)
Study 8 [47]	<ul style="list-style-type: none"> • CNN models and their attributes were studied in this study. This study explores the odd pairing pattern of filters with opposing phases at the bottom layers. The authors introduced concatenated Rectified Linear Unit (CReLU), an effective activation method, and theoretically investigated its CNN reconstruction capability. (Experiment) • The authors introduced CReLU activation to current CNNs. With fewer trainable parameters, they improved CIFAR-10/100 and ImageNet recognition. This study demonstrated that understanding CNNs may improve performance with a simple adjustment. (Results)
Study 9 [48]	<ul style="list-style-type: none"> • The authors' self-calibrated convolution method employs internal communications to expand each convolutional layer's field-of-view, producing richer output features. The self-calibrated convolution technique dynamically establishes spatial and inter-channel interdependence around each place. More data may help CNNs build more discriminative representations. (Experiment) • Using the self-calibrated convolution technique with several backbone models enhanced image recognition, object detection, instance segmentation, and keypoint detection. No network infrastructure changes are needed for these improvements. (Results)
Study 10 [49]	<ul style="list-style-type: none"> • The authors developed a hybrid method to increase CNN accuracy without model retraining. The suggested architectural change is replacing the softmax layer with kNN during inference. Transfer learning often uses this strategy in the neural network's training area. (Experiment) • This study tests the research team's hybrid CNN-kNN architecture on several picture datasets, network topologies, and label noise levels. With noisy labels and huge datasets like CIFAR100, the usual CNN design improves inference accuracy. The authors discovered that using the ℓ_2 norm to neural coding yields statistical advantages. (Results)
Study 11 [50]	<ul style="list-style-type: none"> • The authors examined texture categorization. Machine learning influences texture classification. Since traditional techniques fail at categorization, deep learning may overcome these issues. (Experiment) • CNN and LSTM can pick a robust invariant feature extractor and accurate classifier. Fusion makes categorization rates consistent across data sets, and the proposed model enhances texture classification. The experiments demonstrate CNN-LSTM outperforms SVM and CNN. (Results)
Study 12 [51]	<ul style="list-style-type: none"> • Learning was studied using one-hidden-layer non-overlapping convolutional neural networks with ReLU activation in the research. Research focuses on model estimation. Model parameters are estimated using a minimized non-convex squared training loss function. Since the training set has few Gaussian distribution-generated samples, a good start may accelerate gradient descent. (Experiment) • Despite non-convex learning, the rate is linear. Vanilla gradient descent lags convergence. In contrast to theoretical investigations, infinite samples are expected. Descending approaches may find neural network non-convex learning global optimizer. (Results)
Study 13 [52]	<ul style="list-style-type: none"> • In this study, deep neural networks may replace regular filters, but they still face uncommon cases that cannot be handled by training convolutional filters, a key component of deep learning CNNs. This study aims a complete dual-channel model framework and generalized execution architecture. Also, this study studied how image quality affects a cutting-edge convolutional network. This research examined Gaussian, impulsive, and speckle noise, common in image processing and lossy JPEG compression. The authors have chosen the top three input photo denoising algorithms. (Experiment) • For two channels, the design uses outline-enhanced input. An investigation of small differences between proposed architectural design and other training methods. To generalize the approach, the authors investigated multiple variable-quality sources, including a picture. Authors found dual-channel model outperformed single model. (Results)
Study 14 [53]	<ul style="list-style-type: none"> • This research used CNNs substantially in medical image processing. Due of its disorganized pattern, the pathologist may misidentify or take longer to interpret the image. Performance of deep learning models generally requires strong hardware. Accelerating training and optimizing the histology dataset may prevent these concerns. The authors employed three GTX-1080 GPU configurations to accelerate CNN training. Image processing uses low-pass mean-shift filters. Improving histopathology pictures' unstructured pattern improves image processing. (Experiment) • Image patterns are extracted and refined, boosting clarity. It may uncover and improve unstructured patterns. Prepared for analysis and study. This study examined three GPU performance parameters using 500-epoch training. Batches' sizes were 32, 64, 128 and 256. Mean-shift filtering and 128-batch size may improve convergence. (Results)
Study 15 [54]	<ul style="list-style-type: none"> • This enhanced learning capability may impair the model's ability to generalize to unknown data and induce overfitting. Performance is crucial for training strong CNNs. This work developed LossDA, a dynamically controlled adjustment mechanism that adds a disturbance variable to the fully-connected layer. The size of this variable may be changed, but its trend reflects training loss. The whole regularization procedure may aid CNN generalization. (Experiment) • MNIST was compared to FashionMNIST, CIFAR-10, Cats vs. Dogs, and miniImagenet, often used in academic research, to assess the recommended approach. The empirical data suggest the technique improves the model. This study compares Light and Transfer CNNs InceptionResNet, VGG19, ResNet50, and InceptionV3. The biggest accuracy gain for Light CNNs is 4.62%, F1 3.99%, and Recall 4.69%. Transfer CNNs boost accuracy 4.17%, F1 5.64%, and recall 4.05%. (Results)

- CNN and LSTM fusion to make a robust invariant feature extractor and accurate classifier
- Model parameters estimation using a minimized non-convex squared training loss function
- A complete dual-channel model framework and generalized execution architecture
- Employing three GTX-1080 GPU configurations to accelerate CNN training and a low-pass mean-shift filter
- Exploring a dynamically controlled adjustment mechanism that adds a disturbance variable to the fully-connected layer

4. CONCLUSION

This paper summarized several techniques for improving the performance of Convolutional Neural Networks (CNN), such as; adjusting parameters, image data augmentation, deeper network topology, avoiding overfitting and under-fitting, data normalization, batch normalization, data augmentation, learning rate scheduling, weigh decay, gradient clipping and Adams optimizer. Several studies were mentioned.

In the first research, self-distillation boosts the deepest classifier training by 0.3%–0.7%. Without training convergence, shallow classifiers have flat minimums. Switching training methods may help convergence. Autonomous distillation training beats highly supervised net. Distillation lowers adaptive depth runtime accuracy. Gradients, flat minima, and distinguishing features have been used to study self-distillation. Self-distillation improves compression/acceleration models. Self-distillation may communicate inside one model despite rigorous knowledge transfer investigations. In the second research, low-level and high-level CNN features were segregated by feature size for feature fusion. L and H-Fusion assess feature fusion differently. Multiscale feature fusion is tested with five low-level features of various sizes. CNN performance improves with more auxiliary block low-level properties. In the third research, authors built and tested a transfer deep learning CNN model for chest X-ray COVID-19 detection and classification using various methodologies. Photo preprocessing improves deep learning model input, research shows. Trans-fer learning neutralizes CNN channels' contrast-to-noise ratio, clutter, and colors.

In the fourth research, CNN-RNN-transducer ContextNet may narrow the gap. Squeeze-and-excitation modules provide ContextNet encoder convolution layers global context. Processing and accuracy were balanced by scaling ContextNet widths simple. ContextNet's WER is 2.1%/4.6% with and without LibriSpeech. With 10 M parameters, ContextNet scores 2.9%/7.0% WER on clean/noisy LibriSpeech test sets. Best system performance: 2.0% LM, 4.6% 20M parameters, 3.9%, 11.3%. Larger internal datasets help ContextNet. In fifth research, for nonlinear transformation, multiple activation functions use linear or nonlinear mapping terms. The funnel activation (FReLU) input competed with spatial conditions was proposed. Thus, FReLU may vary nonlinearly. Activating algorithms increases basic competitiveness.

Comprehensive Activation Function competes. PFReLU increases linear/nonlinear mapping space and competition. This study examined four datasets of various sizes. The experimental findings of three well-established methods were investigated. Convolutional neural networks improved in this study.

The sixth research found that automat-ed XCA stenosis detection may identify early health problems. Coronary plaque slowed blood flow and increased MI risk. In big medical imaging datasets, CNNs may detect sick, normal, and differentiated tissues. CNN stenosis detection training improves with HBGM. False picture patches help researchers extend the database and accelerate network convergence. 10,000-photo synthetic dataset containing 50% stenosis and non-stenosis, were used. Statistics employs trustworthy Fréchet Inception Distance. After incorrect dataset training, XCA modifies the network. CBAM self-attention module enhanced network performance. Pre-trained networks employing the suggested generative model outperformed those constructed from scratch. The authors evaluated ResNet topologies with stenosis attention modules. In the seventh research, DELU activation may be positive or detrimental. Fashion Mnist, CIFAR10, and ImageNet experiments were examined. DELU outperformed Leaky ReLU, ReLU, ELU, SELU, Swish, and SERLU. The eighth research analyzed CNN model characteristics. At filters with opposing phases at the lowest layers, an unexpected pairing pattern is studied. Following these observations, the authors suggested concatenated Rectified Linear Unit (CReLU), an effective activation approach, and theoretically examined its CNN reconstruction feature. CNNs' less trainable parameters enhanced CIFAR-10/100 and ImageNet recognition. Research reveals little CNN knowledge changes may improve performance.

The ninth research improves CNN convolutional feature transformation without model changes. Internal communications increase each convolutional layer's field of view, improving output features in the authors' self-calibrated convolution method. Self-calibrated convolution dynamically develops spatial and inter-channel dependency at each point. Convolutional layers may be adjusted without parameters or complexity. Trials show self-calibrated convolution enhances backbone model image, object, instance, and key point recognition. Enhancements do not change network infrastructure. In the tenth research, a hybrid CNN accuracy method without model retraining was described, by using kNN instead of SoftMax for inference. Train neural networks via transfer learning. This research tests the team's hybrid CNN-kNN architecture on picture datasets, network topologies, and label noise. Normal CNNs improve CIFAR100 and noisy label inference.

The eleventh study used texture classification. Deep learning may address classification issues when conventional approaches fail. CNN/LSTM classifiers and extractors work well. SVM and CNN lose against CNN-LSTM. The twelfth used ReLU-activated one-hidden-layer non-overlapping convolutional neural networks for learning. Research focuses on model estimation. Minimized non-convex squared training loss functions can estimate

model parameters. Since the training set has few Gaussian distribution-generated samples, a good start may accelerate gradient descent. Nonconvex linear learning outperforms Vanilla gradient descent in convergence. The paper analyses rapid gradient algorithms. Descending approaches may find neural network non-convex global optimizers. The study addresses gradient-based sample complexity. Convolutional neural networks are investigated using non-smooth ReLU activation.

The thirteenth research suggests that deep neural networks may replace filters in unusual situations that cannot be taught using convolutional filters, a key component of deep learning CNNs. Ambient noise and blur may greatly impact neural network output. Noise's impact on image classification was studied. This research seeks a full dual-channel outline-enhanced input model framework and generalized execution architecture. This research examined how picture quality influences a cutting-edge convolutional network. Lossy JPEG compression was used with Gaussian, impulsive, and speckle noise. The authors selected three top input picture denoising algorithms. Common denoising enhances edges and outlines but eliminates de-tails. Thus, denoised and outline-enhanced photos should not be distorted. Authors generalized the method using variable-quality sources like photos. The two-channel model performed better.

The fourteenth research assessed medical images using CNNs. The histopathologists examine tissue images. Due to its disorganized pattern, the pathologist may misidentify or take longer to analyze the picture. Faster training and histological dataset tweaking may help. Three GTX-1080 GPUs were used to increase CNN training with 500-epoch training. Image processing uses low-pass mean-shift filters. This research depends on extracting and improving visual patterns. Also, it may discover and improve unstructured patterns. 32, 64, 128, and 256 batches' sizes were used, and mean-shift filtering with 128 batches may help convergence.

The fifteenth study demonstrated that CNN input attribute understanding with complicated model topologies is important. Increased learning capacity may restrict model generalization to unknown data and overfit. The literature suggests data augmentation, batch normalization, and dropout enhance generalization. The dynamic adjustment method LossDA uses a disturbance variable to reach the fully-connected layer. Complete regularization may aid CNN generalization. The proposed technique implemented FashionMNIST, CIFAR-10, Cats vs. Dogs, and miniImagenet. Experimental results show this method improved the model. This study compares Light and Transfer CNNs InceptionRes-Net, VGG19, ResNet50, and InceptionV3. Light CNN gives at F1 3.99%, recall 4.69%, and accuracy 4.62%. Transfer CNNs increase accuracy 4.17, F1 5.64, and recall 4.05.

CONFLICT OF INTEREST

Author declares that there is no conflict of interest.

REFERENCES

- [1] Krizhevsky I, Sutskever, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM*. May 2017;60(6):84–90.
- [2] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–8. Jun. 2016.
- [3] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. *International Conference on Machine Learning*, pp. 448–56. 2015.
- [4] Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–7. Jun. 2014.
- [5] Song Z, Liu Y, Song R, Chen Z, Yang J, Zhang C, et al. A sparsity-based stochastic pooling mechanism for deep convolutional neural networks. *Neural Netw*. Sep. 2018;105:340–5.
- [6] Szegedy, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9. Jun. 2015.
- [7] Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult. *IEEE Trans Neural Netw* Mar. 1994;5(2):157–66.
- [8] Forrester NI, Song H, Moskewicz WM, Khalid A, William JD, Kurt K. Squeezenet: alexnet-level accuracy with 50x fewer parameters and 0.5 mb model size. *International Conference on Learning Representations*, 2016.
- [9] Andrew GH, Menglong Z, Chen B, Kalenichenko D, Weijun W, Tobias W, et al. Mobilenets: efficient convolutional neural networks for mobile vision applications. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [10] Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network. In *Advances in Neural Information Processing Systems*. MIT Press; 2014.
- [11] Lei P, Huang Z, Liu G, Wang P, Song W, Mao J, et al. Clinical and computed tomographic (CT) images characteristics in the patients with COVID-19 infection: what should radiologists need to know. *J Xray Sci Technol*. 2020;28(3):369–81.
- [12] Narin, Kaya C, Pamuk Z. Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks, arXiv preprint arXiv:2003.10849. 2020.
- [13] Dai W, Zhang H, Yu J, Xu HJ, Chen H, Luo SP, et al. CT imaging and differential diagnosis of COVID-19. *Can Assoc Radiol J*. 2020;71(2):195–200.
- [14] Elaziz MA, Hosny KM, Salah A, Darwish MM, Lu S, Sahlol AT. New machine learning method for imagebased diagnosis of COVID-19. *PLoS One*. 2020;15(6):e0235187.
- [15] Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Rajendra Acharya U. Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Comput Biol Med*. 2020;121:103792.
- [16] Wang S, Kang B, Ma J, Zeng X, Xiao M, Guo J et al. A deep learning algorithm using CT images to screen for corona virus disease (COVID-19). *Eur Radiol*. 2021;31(8):6096–6104. doi: 10.1101/2020.02.14.20023028.
- [17] Sethy PK, Behera SK. Detection of coronavirus disease (COVID-19) based on deep features, preprints 20200303000. 2020.
- [18] Apostolopoulos ID, Mpesiana TA. Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. *Phys Eng Sci Med*. 2020;43(2):635–640. doi: 10.1007/s13246-020-00865-4.
- [19] Khan AI, Shah JL, Bhat MM. Coronet: a deep neural network for detection and diagnosis of COVID-19 from chest x-ray images. *Comput Methods Programs Biomed*. 2020;196:105581.
- [20] Rahimzadeh M, Attar A. A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2. *Inform Med Unlocked*. 2020;19:100360.
- [21] Zhang L, Song J, Gao A, Chen J, Bao C, Ma K. Be your own teacher: improve the performance of convolutional neural networks via self distillation. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. 2019, pp. 3712–3721.
- [22] Zhang Y, Xiang T, Hospedales TM, Lu H. Deep mutual learning. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4320–8. 2018.
- [23] Krizhevsky A, Hinton G. *Learning Multiple Layers of Features from Tiny Images*. Technical report, Citeseer; 2009.
- [24] Deng J, Dong W, Socher R, Li JL, Kai L, Li FF. Imagenet: a large-scale hierarchical image database. In *Computer Vision and Pattern Recognition*. IEEE; 2009, pp. 248–55.

- [25] Xiaohong Y, Wei L, Yanyan L, Xiaoqiu S, Lin G. Improving the performance of convolutional neural networks by fusing low-level features with different scales in the preceding stage. *IEEE Access*. 2021;9:70273–70285.
- [26] Krizhevsky. *Learning Multiple Layers of Features from Tiny Images*. Toronto, ON, Canada: Univ. Toronto; 2012. pp. 54–7.
- [27] Srivastava RK, Greff K, Schmidhuber J. Training very deep networks. 2015. arXiv:1507.06228. [Online]. Available from: <https://arxiv.org/abs/1507.06228>.
- [28] Pleiss G, Chen D, Huang G, Li T, van der Maaten L, Weinberger KQ. Memory-efficient implementation of DenseNets. 2017. arXiv:1707.06990. [Online]. Available from: <http://arxiv.org/abs/1707.06990>.
- [29] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 2014. arXiv:1409.1556. [Online]. Available from: <https://arxiv.org/abs/1409.1556>.
- [30] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–8. Jun. 2016.
- [31] Lin M, Chen Q, Yan S. Network in network. 2013. arXiv:1312.4400. [Online]. Available from: <https://arxiv.org/abs/1312.4400>.
- [32] Redmon J, Farhadi A. YOLO9000: better, faster, stronger. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6517–25. Jul. 2017.
- [33] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–8. Jun. 2016.
- [34] Heidari M, Mirniaharikandehei S, Khuzani AZ, Danala G, Qiu Y, Zheng B. Improving the performance of CNN to predict the likelihood of COVID-19 using chest X-ray images with preprocessing algorithms. *Int J Med Inform*. December 2020;144:104284.
- [35] Kermany K, Zhang MG. Large dataset of labeled optical coherence tomography (OCT) and chest X-Ray images. *Mendeley Data*. 2018;3. doi: 10.17632/rscbjbr9sj.3.
- [36] Chowdhury MEH, Rahman T, Khandakar A, Mazhar R, Kadir MA, Mahbub ZB, et al. Can AI help in screening viral and COVID-19 pneumonia? arXiv preprint arXiv:2003.13145. 2020. Available from: <https://www.kaggle.com/tawsifurrahman/covid19-radiography-database>.
- [37] Heidari M, Khuzani A, Hollingsworth AB, Danala G, Mirniaharikandehei S, Qiu Y, et al. Prediction of breast cancer risk using a machine learning approach embedded with a locality preserving projection algorithm. *Phys Med Biol*. 2018;63(3):35020.
- [38] Heidari M, Mirniaharikandehei S, Liu W, Hollingsworth AB, Liu H, Zheng B. Development and assessment of a new global mammographic image feature analysis scheme to predict likelihood of malignant cases. *IEEE Trans Med Imaging*. 2020;39(4):1235–44.
- [39] Han W, Zhang Z, Zhang Y, Yu J, Chiu CC, Qin J, et al. *ContextNet: improving convolutional neural networks for automatic speech recognition with global context*. Electrical Engineering and Systems Science; 2020.
- [40] Ying Y, Zhang N, He P, Pen S. Improving convolutional neural networks with competitive activation function. In: *Security and Communication Networks/ Special Issue Machine Learning for Security and Communication Networks*. 2021.
- [41] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 249–56, Italy, January 2010.
- [42] LeCun Y, Bottou L, Bengio Y, Haffner P. Gradientbased learning applied to document recognition. *Proc IEEE*. November 1998;86(11):2278–324.
- [43] Lin M, Chen Q, Yan S. Network in network. *Proceedings of the 2nd International Conference on Learning Representations (ICLR)*, pp. 1–10, Banff, Canada, March 2014.
- [44] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–8, Las Vegas, USA, June 2016.
- [45] Ovalle-Magallanes E, Avina-Cervantes JG, Cruz-Aceves I, Ruiz-Pinales J. Improving convolutional neural network learning based on a hierarchical bezier generative model for stenosis detection in X-ray images. *Comput Methods Program Biomed*. June 2022;219:106767.
- [46] Hu Z, Huang H, Ran Q, Yuan M. Improving convolutional neural network expression via difference exponentially linear units. *J Phys*. 2020, Conference Serie, ICAITA, 1651, 2020.
- [47] Shang W, Sohn K, Almeida D, Lee H. Understanding and improving convolutional neural networks via concatenated rectified linear units. *Proc 33rd Int Conf Mach Learn, PMLR*. 2016;48:2217–25.
- [48] Liu JJ, Hou Q, Cheng MM, Wang C, Feng J. *Improving Convolutional Networks with Self-Calibrated Convolutions*. IEEE CVPR; 2020.
- [49] Gallego AJ, Pertusa A, Calvo-Zaragoza J. Improving convolutional neural networks' accuracy in noisy environments using k-nearest neighbors. *Appl Sci*. 2018;8:2086.
- [50] Rao MS, Reddy BE. An improved convolutional neural network with LSTM approach for texture classification. *Int J Emerg Trends Eng Res*. July 2020;8(7):3827–33.
- [51] Zhang S, Wang M, Xiong J, Liu S, Chen PY. Improved linear convergence of training CNNs with generalizability guarantees: a one-hidden-layer case. *IEEE Trans Neural Netw Learn Syst*. June 2021;32(6):2622–2635.
- [52] Yim J, Sohn KA. Enhancing the performance of convolutional neural networks on quality degraded dataset. *International Conference on Digital Image Computing: techniques and Applications*, 2017.
- [53] Haryanto T, Suhartanto H, Murni A, Kusmardi K. Strategies to improve performance of convolutional neural network on histopathological images classification. *2019 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, 2019.
- [54] Liu J, Zhao Y. Improved generalization performance of convolutional neural networks with LossDA. *Appl Intell*. 2023;53:13852–6.