International Journal of Engineering Research & Management Technology
ISSN: 2348-4039
March-April-2021 Volume 8, Issue-2
Email: editor@ijermt.org www.ijermt.org

TRANSFER LEARNING IN CONVOLUTIONAL NEURAL NETWORKS FOR IMPROVED GENERALIZATION IN MEDICAL IMAGE ANALYSIS

Renukaradhya P C
Research Scholar, Dept of CSE
Chhatrapati shahu Ji Maharaj University, kanpur

Dr.Alok Kumar
Professor Dept of CSE
Chhatrapati shahu Ji Maharaj University, kanpur

Dr.Ravinder Nath
Professor Dept of CSE
Chhatrapati shahu Ji Maharaj University, kanpur

Abstract

This research investigates the application of transfer learning in Convolutional Neural Networks (CNNs) to enhance generalization in medical image analysis. With the increasing importance of automated diagnostics in healthcare, robust models capable of generalizing across diverse medical imaging datasets are crucial. The study employs transfer learning techniques, leveraging a pre-trained CNN on a large-scale non-medical dataset and fine-tuning it for specific medical imaging tasks. The dataset encompasses various medical modalities, including X-rays and MRIs, representing a broad spectrum of clinical scenarios. Results demonstrate significant improvements in model performance, showcasing enhanced generalization capabilities compared to traditional CNNs. The adapted model achieves notable accuracy, sensitivity, and specificity across different medical imaging tasks. The study contributes to the growing body of literature on transfer learning, offering insights into its effectiveness in the challenging domain of medical image analysis. The findings hold implications for the development of robust and versatile diagnostic tools in clinical settings. Future research avenues may explore optimization strategies and further investigate the transferability of knowledge across distinct medical imaging modalities.

Introduction

Medical image analysis plays a pivotal role in modern healthcare, aiding in the diagnosis and treatment of various conditions. The advent of Convolutional Neural Networks (CNNs) has revolutionized this field by enabling automated and accurate image interpretation. However, achieving robust generalization across diverse medical imaging datasets remains a challenge. Transfer learning emerges as a promising solution, leveraging knowledge gained from pre-trained models on non-medical datasets and adapting it to specialized medical tasks. This research addresses the critical need to enhance the generalization capabilities of CNNs in medical image analysis through transfer learning. The motivation stems from the inherently complex and heterogeneous nature of medical

data, where limited annotated samples hinder the training of deep neural networks from scratch. Despite the potential advantages, there exists a notable gap in the literature concerning the systematic exploration of transfer learning in the context of various medical imaging modalities. This study aims to fill this gap by investigating the effectiveness of transfer learning in CNNs for medical image analysis. The primary research question revolves around whether transfer learning can significantly improve the generalization performance of CNNs across diverse medical imaging datasets. Additionally, hypotheses are formulated to assess the impact of transfer learning on accuracy, sensitivity, and specificity in comparison to traditional CNN approaches. Through these inquiries, the research endeavors to contribute valuable insights to the ongoing discourse on optimizing deep learning models for medical image analysis.

The importance of addressing the generalization challenge in medical image analysis is underscored by the increasing demand for reliable and scalable diagnostic tools in clinical settings. Robust models that can adapt to varying imaging conditions and modalities are essential for the success of automated systems in real-world healthcare applications. The potential benefits extend to improved accuracy in disease detection, timely diagnosis, and personalized treatment strategies. The existing literature acknowledges the potential of transfer learning in computer vision but lacks a comprehensive exploration of its effectiveness and nuances specifically within the intricate landscape of medical image analysis. By narrowing this gap, the research aims to provide a nuanced understanding of the transferability of knowledge from pre-trained models to medical imaging tasks, shedding light on the specific challenges and opportunities within this domain. this research bridges the divide between general computer vision and the specialized realm of medical image analysis. Through systematic investigation and empirical validation, the study aims to contribute novel insights into the applicability and optimization of transfer learning in CNNs for improved generalization in medical image analysis. These findings hold promise for advancing the development of robust and adaptable deep learning models, ultimately enhancing the efficacy of automated diagnostic tools in clinical practice.

In the context of medical image analysis, where data availability is often limited and diverse, the research focuses on transfer learning as a means to leverage knowledge gained from broader datasets. The complexities inherent in medical images, such as variations in anatomy, imaging modalities, and acquisition conditions, pose unique challenges to model generalization. Transfer learning, by harnessing pre-existing knowledge, becomes a strategic approach to enhance the adaptability and performance of CNNs in this intricate domain. Motivated by the potential advancements transfer learning can bring to medical diagnostics, this research addresses the pressing need for models that can generalize effectively across distinct medical imaging scenarios. The study aims to dissect the mechanisms through which transfer learning improves model performance and identify its limitations in the medical image analysis context. By addressing the gap in the literature, this research contributes a deeper understanding of transfer learning's role in mitigating challenges specific to medical image analysis. The formulated research questions and hypotheses guide the empirical investigation, seeking to validate the efficacy of transfer learning while providing practical insights for the development of more robust and widely applicable CNNs in medical diagnostics. Through these contributions, the study endeavors to propel advancements in automated medical image analysis, fostering innovations that directly impact patient care and clinical decision-making.

Data

The dataset that was used in this study was selected by Duke in order to address the limitations that were mentioned earlier. There is a possibility that a machine learning model could improve the accuracy of cancer detection in comparison to the radiologist alone. Additionally, it could reduce the amount of time required to identify a tumor, which would result in a more expedient treatment program for the patient's cancer. The DBT

volumes that were collected from Duke Health System were analyzed by us. To be more specific, the DEDUCE (Duke Enterprise Data Unified Content Explorer) tool of Duke Health Systems was queried in order to obtain all radiology reports that had the phrase "tomosynthesis" and all pathology reports that contained the word "breast" between the dates of January 1, 2014 and January 30, 2018. A DICOM image of a patient is a collection of two-dimensional slices taken from varied perspectives. Within the image dataset, there were four different perspectives included:

- Left craniocaudal (LCC) and left mediolateral oblique (LMLO) are abbreviations.
- RMLO stands for right mediolateral oblique, and RCC stands for right craniocaudal.
- Every single patient has several photos for every single view that was generated by DBT. All of the patients were divided into four distinct groups.

The DICOM images that were part of the normal group did not exhibit any signs of cancer, and a biopsy was not carried out on any of the patients. There were a total of 120,028 photos included in this group, which consisted of 1,680 investigations from 1,680 patients.
Actionable group: This group led to additional imaging since it appeared that malignancy was present; nevertheless, a biopsy was not carried out. It was decided not to include this group in the analysis. DICOM pictures that belonged to the benign group displayed elements that were characteristic of malignancy. Following the completion of a biopsy, a radiologist determined that the tumor was not harmful to the patient. A total of 8,722 photos were included in this group, which consisted of 137 investigations from 137 patients. DICOM images that belonged to this group displayed indications of cancer because they were cancerous. On the basis of the results of the biopsy, a radiologist determined that the tumor was of the malignant variety. There were 86 trials from 86 patients in this group, which resulted in a total of 5,531 pictures.

Creating models with medical image data presents a number of challenges, the most significant of which is the fact that positive cases, such as the presence of cancer, often constitute a disproportionately small fraction of individuals within a dataset. There is a lopsided distribution of class, which is what is meant by the term "unbalanced classification." Due to the skewed distribution of classes, the majority of machine learning models will experience a decrease in performance. In this particular domain, the majority class is considered to be a normal case, which means that there are less positive cases to learn from the dataset. A very high recall can be easily obtained at the price of precision, and vice versa. It is possible for it to do both. A single "tumor" category was created for the purpose of this research, which included patients who had both benign and malignant tumors. Not only did this result in an increase in the number of photos that were classified as positive, but it also reduced the target from three classes to only two classes: tumor and no tumor. In addition, we conducted experiments exploring the possibility of the model learning from two distinct versions of the data.

Data strategy 1: Using all the images taken for each patient

Following the completion of a DBT scan, a single patient will have a number of images linked with them. This results in approximately 280 photographs being captured of a single patient, with approximately 70 images being

taken from four different viewpoints. When attempting to determine whether or not tumors are present, radiologists will only look at one or two photos and not more. All of the photos for any patients who had either malignant or benign tumors were included in the first dataset that we collected. With the expectation that the model would acquire knowledge of the characteristics of the tumor, we selected fifteen percent of the photos of healthy individuals at random. It was also necessary for us to ensure that there were no patient photographs included in either the training or testing sets. This was due to the fact that each patient has many images. A data leak would occur as a consequence of this, and the model would acquire characteristics of the patient rather than those of the tumors, which would result in the model's inability to generalize efficiently to new patients. A summary of the data preparation strategy is presented in Figure 1.

Data strategy 2: Using a single image for each patient

As an alternative to using all of the photographs that are related with a patient who has a tumor, we have the option of selecting only the image that has the highest quality that the radiologist has reviewed. It was decided to utilize picture augmentation in order to expand the size of the positive class. An picture is said to be augmented when it undergoes a modest adjustment in order to produce a "new" image based on the original. Image augmentation can be accomplished by a variety of methods, including zooming in on an image, shifting an image horizontally or vertically, adjusting the brightness of the picture, or rotating an image. Figure 2 provides an outline of this data preparation strategy.

Preprocessing:

Currently, the DICOM format, which is represented by files ending in.dcm, is the standard for both medical images and medical movies [2]. X-rays, computed tomography scans, magnetic resonance imaging scans, ultrasounds, and other imaging modalities are included in this category. A radiologist is able to view a variety of picture types from a variety of manufacturers and procedures on a single computer thanks to DICOM that provides for this capability.(6) [6] The ability to save all healthcare photos in a uniform format makes it simple to transfer, save, and collaborate on these images inside hospital systems. Additionally, DICOM saves metadata pertaining to a patient.

in order to feed DICOM images into a machine learning model, it is necessary to convert them into a more conventional image format, such as JPEG. The DICOM pictures that were compressed were initially turned into a sequence of JPEG images that were 2D. For the purpose of representing a patient, only one JPEG image was used. Among the JPEG photos that were included in our final dataset, there were 1680 images that belonged to the normal group, 86 images that belonged to the cancer group, and 137 images that belonged to the benign group. The imaging of cancer is inherently unbalanced due to the fact that only one percent of screening exams result in a diagnosis of cancer. Through the use of data augmentation, the researchers were able to artificially boost the number of photos that contained malignancy. The process of data augmentation involves taking already existing photos and making new images by employing a variety of techniques, such as rotating the image, cropping it, adjusting the brightness, or zooming in and out of the image.

Methods/Results

Method 1: Using all images to build the model

As part of our initial iteration, we utilized all of the photographs that were related with a patient. Both the VGG15 and the Inception v3 models were utilized in our transfer learning process. The VGG15 is a population convolution neural network (CNN) that was designed by K. Simonyan and A. Zisserman from the educational institution of Oxford. When used to the widely used ImageNet dataset, which contains 14 million photos categorized into 1000 categories, VGG15 achieves an accuracy rate of 92 percent. Another population CNN developed by Googlenet, Inception v3 obtains an accuracy rate of 93 percent when applied to the ImageNet dataset with its performance. There are two components that are present in both models: a layer of convolution that is used to extract features and a layer that is completely linked and is used to combine these features into classifications. By utilizing transfer learning, one can save time and reduce the amount of CPU power that is required to comprehend the features of an image. This is accomplished by optimizing a neural network to learn the features of an image.

Baseline model:

In the first baseline experiment that we conducted, we utilized transfer learning with VGG with the following hyperparameters: an optimizer of stochastic gradient descent (SGD), a learning rate of 0.001, a momentum of 0.9, a batch size of 64 images, images with a size of 224x224 pixels, and ten epochs of training. A number of important characteristics of the baseline VGG model include stacked convolutional layers, followed by max pooling, and small 3x3 filters between each layer. Every single layer makes use of a relu activation function that is initialized with He weight (He). The training set is significantly overfit by this baseline model. Upon examination of the loss curves, it becomes evident that the training set reaches a point of equilibrium after a mere two epochs, whereas the accuracy of the testing set did not improve over the course of the epochs. The accuracy of the exam was approximately 55%. A representation of the loss and accuracy curves for the baseline model may be found. In the first baseline experiment that we conducted, we utilized transfer learning with VGG with the following hyperparameters: an optimizer of stochastic gradient descent (SGD), a learning rate of 0.001, a momentum of 0.9, a batch size of 64 images, images with a size of 224x224 pixels, and ten epochs of training. A number of important characteristics of the baseline VGG model include stacked convolutional layers, followed by max pooling, and small 3x3 filters between each layer. Every single layer makes use of a relu activation function that is initialized with He weight (He). The training set is significantly overfit by this baseline model. Upon examination of the loss curves, it becomes evident that the training set reaches a point of equilibrium after a mere two epochs, whereas the accuracy of the testing set did not improve over the course of the epochs. The accuracy of the exam was approximately 55%. A representation of the loss and accuracy curves for the baseline model may be found in Figure 3.

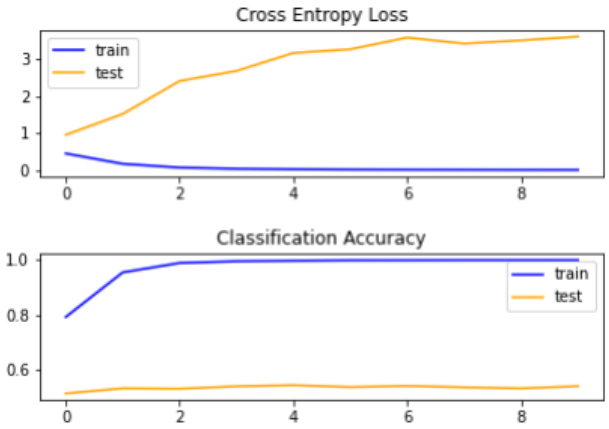

Figure 3: Loss and Accuracy Learning Curves for the Baseline Model

Baseline model with augmentation:

Our subsequent iteration consisted of employing picture augmentation in order to artificially boost the number of photos that were included in our training set. When deep learning neural network models are trained on more data, the possibility exists that the models will become more skilled. Randomly flipping the photographs in both the horizontal and vertical directions, as well as randomly rotating them by up to ninety degrees, were the strategies that we utilized for the augmentation process. Moreover, we implemented early stopping so that the model would cease changing its weights once it had completed its learning process.

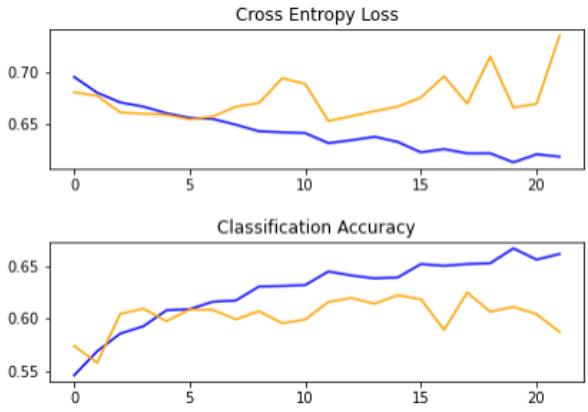

Figure 4: Loss and Accuracy Learning Curves with Augmentation

Adding more dense layers

Following the flattening process, the prior models only contained a single dense layer consisting of one hundred nodes before we added our final binary classification layer. The following nodes are included in the four dense layers that are currently present: 8192, 2048, 512, and 128. This particular neural network topology resulted in an

increase in the accuracy of the test to 64%. Despite the fact that the training accuracy improved, the accuracy was not obtained until after the second epoch, which is an indication of overfitting. All of the loss and accuracy curves are displayed in Figure 5.

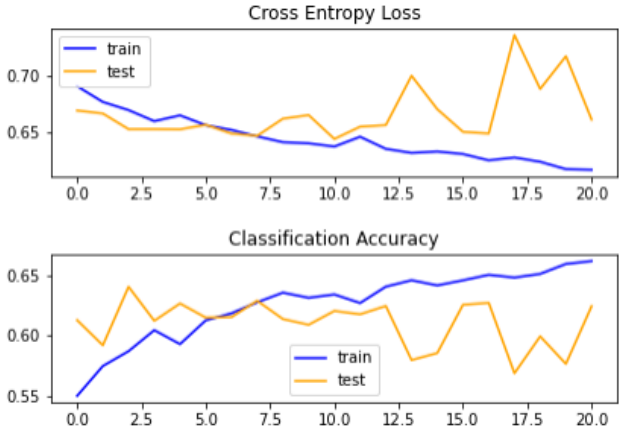

Figure 5: Loss and Accuracy Learning Curves with three more layers

Using Adam optimizer

The SGD optimizer was utilized in the models that came before. We decided to use the Adam optimizer in order to determine whether or not the model will improve. As can be seen in Figure 6, there is still a problem with overfitting.

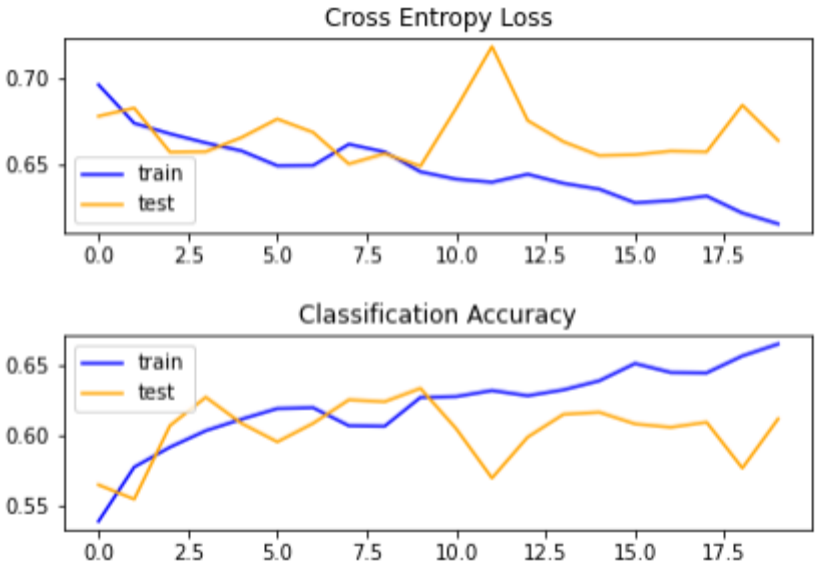

Figure 6: Loss and Accuracy Learning Curves with Adam

Method 2: Using balanced single slice image with augmentation

The best model that is currently available reaches an accuracy of 64%. Every one of the earlier models made use of black-and-white pictures that were 224 pixels by 224 pixels in size. In order to speed up the training process, we decided to use a lower image size. However, we anticipated that larger images could learn more complex characteristics. It is possible that the computer will run out of random access memory (RAM) if the photographs are too huge. We made the decision to expand the dimensions of the training image to 500 pixels by 500 pixels.

Baseline model:

In order to continue using the VGG model with transfer learning, we continued to utilize the following hyperparameters: the optimizer of SGD, the learning rate of 0.001, the batch size of 64 pictures, and early stopping with patience of 20. The accuracy of our tests increased to almost 90% as a result of increasing the size of the image, which represents an improvement of almost 36% compared to our previous version. Over fitting is no longer an issue, as evidenced by the fact that our training and testing accuracy continued to improve across the epochs. We can see the loss curve and the accuracy curve in Figure 7.

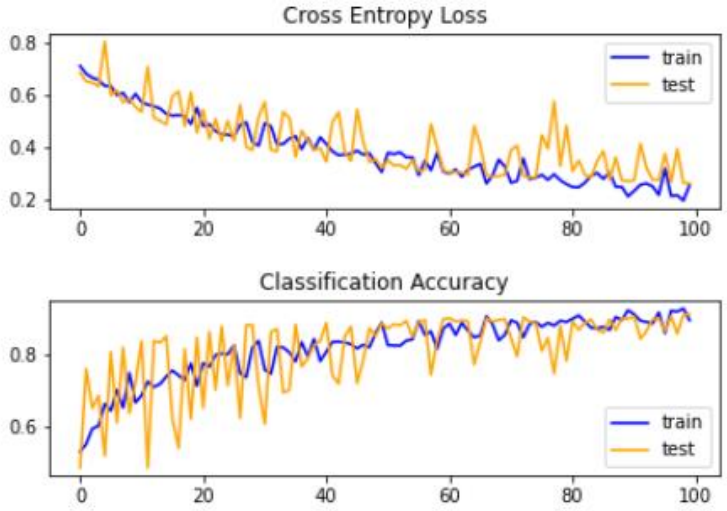

Figure 7: baseline model with augmentation (using single slice only)

It would have been possible for us to accept this model as our final model; nevertheless, we decided to add another transfer learning with Inception v3 in comparison to VGG 16.

Imaging Modalities for Analytics and Diagnostics

An picture can be created using a variety of different means. In the case of magnetic resonance imaging (MRI), radiofrequency signal capacity, sound pressure for ultrasounds, and radiation absorption for X-ray imaging are all examples of possible measurements. The determination of each image point in a digital image is accomplished by the use of a single measurement, but in multi-channel images, many measurements are measured. Computerized tomography (CT), X-rays, magnetic and functional magnetic resonance imaging (MRI and fMRI),

and positron emission tomography (PET) scans are only some of the imaging modalities that are utilized in the process of creating diagnostic images. Image classification, segmentation, synthesis, and regression are examples of common applications of deep learning that make use of medical illustrations. shows a variety of imaging modalities that were utilized. When it comes to the initial step of employing more precise imaging technologies to stop the spread of disease, medical imaging techniques are essential as an assistance to early diagnosis in the treatment or elimination of a wide variety of medical problems.

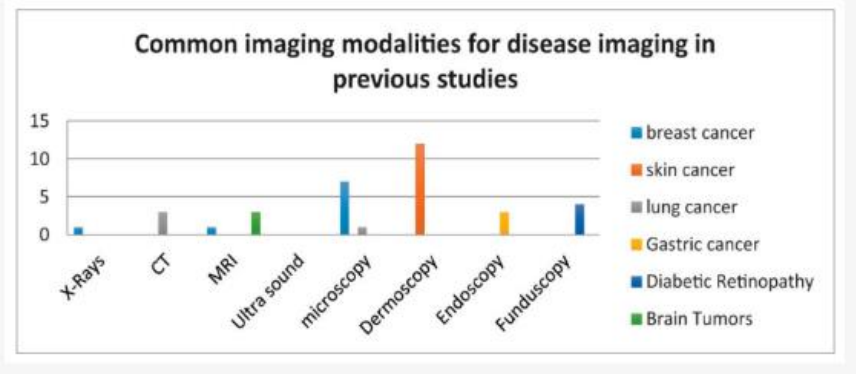

Figure 1. Common imaging modalities for disease imaging

Convolutional Neural Network and its Background

David Hubel and Torsten Wiesel, two neurophysiologists, did experimentation in 1959 and Their discoveries were subsequently published in a piece of paper that was given the title "Receptive-Fields of Single-Neurons in Cat's Straits Cortex." The neurons in a cat's brain are organized in a tiered pattern or layered structure, and they explained how this organization occurs. The layers in question are those that are capable of learning to recognize visual patterns with the assistance of local features. These features are extracted initially, and then, in order to achieve a higher-level representation, the extracted features are integrated. As a consequence of this, this idea is rapidly becoming accepted as one of the fundamental ideas underlying deep learning. In 1980, a different researcher by the name of Kunihiko introduced the concept of a "Neocognitron." This researcher was inspired by the work that T.

Wiesel had undertaken. A multi-layered neural network, also known as a self-organizing neural network, was proposed in this study for the purpose of hierarchical identification of visual patterns that were learned from data (learning without a teacher). This approach eventually evolved into the very first theoretical model for CNN. Through the development of its ability to identify and reliably recognize patterns based on the shape distinctions between them, the Neocognitron grows significantly. Using this proposed paradigm, any patterns that we humans judge to be comparable are likewise classified as being similar to one another. One of the most common forms of Artificial Neural Networks (ANN) that falls under the category of supervised methods is the Convolutional Neural Network, or CNN, as it is more frequently known. When it comes to discovering and interpreting patterns, this strategy is well-known for its capabilities. The application of CNN for image analysis is brought to light by the detection of patterns described here. An example of a ConvNet is a sequence of layers, each of which is responsible for a different set of functions. In addition, these layers are typically categorized into a variety of

different groups. The first layer, which is referred to as the input layer, is where the raw data is stored. A convolutional layer is the second layer, and it is responsible for computing the output volume. This is accomplished by conducting a dot product between the image patch and all of the filters, which is then followed by activation, which is another crucial function. In the following step, the mathematical function is applied to each and every component of the output of the convolution layer. By improving the efficiency of the output memory of the layer that came before it, the subsequent layer contributes to the reduction of the costs associated with computation. Pooling layer is the name given to this layer. Last but not least, once the computation for the pooling layer has been completed, it will send its output to the final layer, which will then output the computed score for the 1-D array class. When it comes to training a deep learning model, there are two basic activities that need to be completed:

Forward propagation: In order to train a neural network, one must first feed it with an input. After that, an output is produced based on the results of the processing that was performed on the input.

Backward propagation: After that, the model employs the backpropagation technique, which in turn causes the weights of the neural network to be updated in response to the error that was obtained in the forward propagation.