A Machine Learning Application for Medical Image Analysis Using Deep Convolutional Neural Networks (CNNs) and Transfer Learning Models for Pneumonia Detection

Abstract: Pneumonia is still a major problem in world health, causing a lot of illness and death, particularly among the most susceptible people. For care and therapy to be successful, a prompt and precise diagnosis is essential. Recent years have seen encouraging outcomes in medical image processing tasks when using deep learning methods, including Convolutional Neural Networks (CNNs). Using deep convolutional neural networks (CNNs) and transfer learning models, we provide a machine learning application for detecting pneumonia in medical photographs. Our suggested system makes use of convolutional neural network (CNN) designs that have already been trained using massive natural picture datasets, including VGG, ResNet, and DenseNet. We hope that by training these algorithms on a collection of chest X-ray pictures, we may modify their characteristics to better diagnose pneumonia. Furthermore, we investigate data augmentation methods to improve the models' generalizability, especially in cases when there is a lack of labelled data. Standard measures including precision, sensitivity, specificity, and area under the curve of the receiver operating characteristic (AUC-ROC) are used to assess the performance of the created models on a heterogeneous dataset of chest X-ray images.

Keywords: Pneumonia detection, Medical image analysis, Deep learning, Convolutional Neural Networks (CNNs), Transfer learning, VGG, ResNet, DenseNet, Chest X-ray images, Data augmentation, Accuracy, Sensitivity, Specificity, AUC-ROC

Introduction

Numerous sectors have made use of AI, due to the technology's meteoric rise in popularity in recent years. Including industries like banking, data mining, and manufacturing. Artificial intelligence (AI) can help humans with a wide variety of activities if we train it correctly, at one point required intricate human involvement. The use of AI in healthcare stands out among the most significant applications. Pneumonia is one of the most prevalent illnesses globally, affecting an estimated 450
million people annually. According to Nair and Niederman (2011), pneumonia was the seventh most common cause of death in the US in 2009, indicating that a significant number of individuals are affected by this disease. Chest X-ray is a commonly used and proven method for detecting pneumonia, and modern AI offers a range of approaches for accurate picture categorization. As an alternative, convolutional neural networks are thought to be the best deep learning method for handling picture categorization issues[1][2]. Building a deep learning model that can identify pneumonia using convolutional neural networks is the focus of this thesis. Its purpose is to provide a reliable way for individuals to determine whether they have pneumonia in the event that there is a scarcity of medical professionals or necessary medical equipment. Several research have focused on the identification and classification of chest radiographs. Computer vision and deep learning (DL) have made great strides in medical identification, especially for chest X-ray images of pneumonia. In 2017, for example, Rajpurkar et al. built a model and developed an algorithm named CheXNet to correctly detect thirteen distinct chest diseases, including pneumonia, using chest X-ray images[3][4]. By integrating a new decision-support mechanism with the VGG16 model, this model enhances the ability to differentiate between viral and bacterial pneumonia in children's chest X-ray images. Data analysis and prediction are at the heart of machine learning, a subfield of artificial intelligence. To summarise, the goal of machine learning is to understand and use human-comprehensible models by deciphering data structures [1]. Although it falls within the umbrella of computer science, machine learning stands apart from traditional approaches to computing. If we talk about computers doing computations or solving problems by following established sets of instructions, we're talking about conventional algorithmic computing. Contrarily, computers are able to learn given data inputs and provide outputs that are scientifically inside a particular range thanks to machine learning algorithms. Machine learning's value lies in the fact that it enables computers to autonomously generate models from data samples, allowing them to make judgements based on incoming data [1]. Many different kinds of work fall under the umbrella of machine learning. The techniques utilised to accept learning or deliver feedback to the created system are the basis for these groupings [2]. When it comes to machine learning, supervised and unsupervised learning are two of the most used methods. Algorithms are trained in supervised learning using examples of input and output data that have been labelled by humans, while algorithms are trained in unsupervised learning using data that does not have labels [2][5].

**Supervised Learning**

In supervised machine learning, algorithms are trained using labelled data sets. Repeatedly adjusting the weights on the labelled input set allows the computer to learn and acquire the appropriate output for prediction. The two main categories of supervised machine learning are two primary groups: classification (using decision trees or a random forest) and regression (using linear or logistic regression)[6][7]. What follows is a description of neural networks as an algorithm for supervised learning, which is the main topic of this thesis. Among the many supervised machine learning applications, some of the best are text categorization, picture identification, price prediction, and email spam detection[8][9].

**Unsupervised Learning**

To train, unsupervised machine learning makes use of datasets that are not labelled. In order to create clusters that match the patterns seen in the unlabeled datasets, the computer use algorithms to examine the links between them[10][11]. By itself, unsupervised ML may learn about the data's commonalities and contrasts. There are two primary types of unsupervised machine learning: clustering and association. When it comes to data exploration, market segmentation, picture compression, and market basket analysis suggestion, unsupervised machine learning is the way to go[12].

**Semi-Supervised Learning**

Incorporating semi-supervised learning techniques into the proposed framework for pneumonia detection enhances the model's ability to utilize unlabeled data effectively, thereby potentially improving performance, especially when labeled data is limited or expensive to acquire[13][14]. Semi-supervised learning algorithms make use of both labeled and unlabeled data during training. In the context of pneumonia detection, where obtaining large amounts of labeled chest X-ray images might be challenging due to the need for expert annotation, leveraging unlabeled data becomes crucial[15][16]. One common approach in semi-supervised learning is to use a combination of supervised and unsupervised learning methods, where the labelled data is used to train the model, and the unlabeled data is used to improve the model's performance. This approach can be particularly useful in scenarios where obtaining labelled data is difficult or expensive.
learning is to use a combination of labeled and unlabeled data to jointly optimize the model's parameters[17]. This can be achieved through techniques such as consistency regularization, where the model is encouraged to produce consistent predictions for unlabeled data under different perturbations or transformations[18]. By doing so, the model learns to generalize better and becomes more robust to variations in the data distribution. Another strategy is self-training, where the model iteratively generates pseudo-labels for unlabeled data based on its current predictions and then incorporates these pseudo-labeled samples into the training set. This process can help the model refine its decision boundaries and improve its performance over time[19].

In the proposed pneumonia detection framework, semi-supervised learning can be applied by augmenting the labeled dataset with a larger pool of unlabeled chest X-ray images. The model is then trained on both labeled and unlabeled data using appropriate semi-supervised learning algorithms. By leveraging the additional unlabeled data, the model can learn more discriminative features and potentially achieve higher performance in pneumonia detection tasks[20].

Evaluation of semi-supervised learning techniques within the framework would involve comparing the performance of models trained with labeled data only against those trained with both labeled and unlabeled data. Metrics such as accuracy, sensitivity, specificity, and AUC-ROC would be used to assess the effectiveness of semi-supervised learning in improving the model's performance[21].

**Reinforcement Learning**

An agent learns to navigate its surroundings in a way that maximises cumulative rewards using a machine learning paradigm called Reinforcement Learning (RL). The agent gains knowledge by experimenting with different activities in its surroundings and obtaining positive or negative feedback. As RL progresses, the objective is to identify the best policy—a mapping between states and actions—that maximises the projected cumulative reward over time.

**Deep Learning**

One kind of machine learning, deep learning (DL) uses a network architecture similar to the human brain, complete with many layers of neural connections. By using deep learning, several computer processing layers are able to acquire knowledge from training data. By fine-tuning the training layers and modifying the weights, the deep learning model is derived at the end. In particular, the neural network uses its node layers to transmit data. With each iteration of the data-passing process, the preceding layer adds to and refines the data that is passed on to the subsequent node. There is no hard and fast rule about how many training layers a model should have, although in practice, using more than one often yields better and more accurate results. When training using deep learning, users are not required to build data features, unlike with classical machine learning. During training, it picks up features and then comes up with its own[22]. Even simpler tasks like route prediction and cargo recognition are beyond the capabilities of machine learning. On the other hand, deep learning is better suited to complicated jobs, like autonomous technology[23]. Deep learning outperforms classical machine learning when it comes to training models, but it does it at the expense of computing resources, particularly GPUs. Because of the intricate layer structure and methods, training often takes more time than machine learning[24].

**Convolutional Neural Network**

One of the neural networks used in deep learning for picture identification and classification applications is the convolutional neural network (CNN). By training using biased and weighted inputs and then extracting the best features, convolutional neural networks are able to differentiate between pictures. One major benefit of convolutional neural networks is that they may work directly on the raw picture without requiring any preprocessing, unlike previous algorithms or neural networks that needed feature extraction and other forms of preprocessing. Among the many present uses for convolutional neural networks, two stand out: pictures (for tasks such as picture identification and classification) and voice (for tasks such as natural language processing and speech recognition).[7][8][9]

**Pneumonia**
Inflammation of the air sacs in the lungs occurs as a result of the illness known as pneumonia. Pus or fluid may collect in the air sacs, leading to a purulent cough, chills, fever, and trouble breathing\cite{12}. Pneumonia may be caused by a wide range of microorganisms, such as bacteria, viruses, and fungus. There is a wide severity spectrum for pneumonia, from moderate to potentially fatal\cite{25}. Pneumonia is a dangerous illness that disproportionately affects newborns, young children, the elderly, and those with compromised immune systems. Each year, pneumonia claims the lives of over 50,000 Americans and sends roughly 1 million individuals to the hospital for treatment\cite{6}. Classification of the main forms of pneumonia is based on the following criteria: the source of infection, the mode of transmission, and the method of acquisition.

Streptococcus pneumonia is the leading cause of bacterial pneumonia. Bacterial pneumonia may also be caused by Chlamydia pneumonia and Legionella pneumophila.

Respiratory viruses often cause pneumonia, particularly in younger children and the elderly.

When compared to bacterial pneumonia, viral pneumonia is often less severe and has a shorter duration. Both the physical exam and a chest X-ray may confirm pneumonia, but an x-ray exam will provide your doctor a better look at your lungs, heart, and blood vessels, which can help them diagnose the disease more accurately, especially if your symptoms are severe and consequences are a concern. A radiologist looking at an x-ray of the lungs will be on the lookout for infiltrates, which are white patches, which indicate an infection. Abscesses and pleural effusions (fluid around the lungs) are two pneumonia complications that may be detected during this exam\cite{7}.

Although the majority of child fatalities are caused by pneumonia in South Asia and Africa, it remains the top cause of death for children under the age of five worldwide. Despite the fact that measles, AIDS, and malaria have received greater focus in recent years, pneumonia remains the leading cause of death among children under the age of five\cite{9}.

When the lungs get infected with pneumonia, it is known as an acute respiratory tract infection (ARTI). Pneumonia causes a buildup of pus and fluid in the alveoli of the lungs, which decreases oxygen intake and makes breathing difficult. Bacteria and viruses are among the several potential causes of pneumonia\cite{9}. When a person breathes in, air travels to their alveoli via their airways and alveolar ducts. The alveoli are tiny air sacs with blood-filled walls. This is the location where blood and air exchange gases, namely carbon dioxide and oxygen\cite{10}. Important to the result is the body's reaction to infection, which manifests as pus building up in the lungs. The pus comprises several components, including blood, white blood cells (especially neutrophils), and plasma proteins (especially opsonins). In order to defeat infection, certain cells and proteins must be present. For this reason, it is crucial to transport these cells and proteins to the lungs, where the bacteria reside, in order to prevent pneumonia.

Nevertheless, this very pus poses a threat. Despite their usefulness in destroying bacteria, neutrophils may cause lung harm with their toxic and derivative products. One example is the production of hypochlorite, the active ingredient in bleach, by neutrophils in pneumatic lungs. While this is beneficial for eliminating germs, it is detrimental to lung cells. Pulmonary edema, a buildup of fluid in the lungs, is another consequence of plasma protein accumulation. As a result of pulmonary edema, it becomes more difficult to breathe, which in turn hinders the body's ability to use oxygen and carbon dioxide, two gases that are essential for proper bodily function. Keeping this pus buildup under control is essential for fighting off germs without causing respiratory issues\cite{11}. Homophilies influenza type b (Hib), Streptococcus pneumonia, and respiratory syncytial virus (RSV) are the most prevalent infections. People at highest risk for pneumonia include those who are under the age of five, those who are 65 or older, and those with preexisting medical conditions. S. pneumonia is the most common bacterial pneumonia in children under the age of five in developing countries. Hib is the second most common bacterial pneumonia in children, and RSV is the most common viral pneumonia in children under the age of two. Pneumococcal infections are more common in those who already have preexisting medical issues, such as sickle cell anaemia, chronic liver, lung, or heart illness. Another group that is more vulnerable to contracting this illness is those who are HIV positive, have had an organ transplant, or are taking any medicine that lowers their immune system's ability to fight off infections\cite{6}.

A person's pneumonia status might be ascertained by the initiative. We have successfully trained, tested, and validated using chest X-ray pictures collected from Kaggle. When it comes to training and certification, the
CNN model is where it's at. Everything went well with the picture resizing, scaling, and capturing. Processing in batches using keras models. After that, we checked the model's ability to provide: Multiple input pictures bound. which reveals to Mark that image as either "Pneumonia" or "Normal" We have finished the first phase of the project's training and testing.

**Dataset Description**

Fig 1: Images with Pneumonia

![Fig 1: Images with Pneumonia](image1.jpg)

Fig 2: images without Pneumonia

![Fig 2: images without Pneumonia](image2.jpg)

The dataset includes filtered chest X-ray pictures from retrospective cohorts and is sourced from the Kaggle platform [8]. The chest X-ray pictures are in RGB format and show the front and back of the patient. The dataset consists of over 5,900 JPEG X-ray pictures that are included in the Binary Classification (Pneumonia/Normal) and is organised into three folders: Train, Test, and Validation. The data is adjusted to enhance overall accuracy and efficiency; the training set contains around 5200 photos and the validation set 600 images.

**Data Augmentation**

An artificially larger dataset may be created by the process of data augmentation, which involves extracting new data points from preexisting datasets. Data augmentation might include tweaks to existing data or the use of machine learning models to generate additional data points inside the original data's latent space. In addition to lowering operational expenditures associated with data gathering, data augmentation aids in enhancing model performance. Nearly all state-of-the-art deep learning applications use data augmentation methods. These include semantic segmentation, object identification, picture categorization, image recognition, natural language understanding, and many more. Augmented data enhances deep learning model performance and outputs by offering more diverse and extra instances for training datasets. In order to help the Deep Learning model pick up on all the nuances in the training images, we undertake a number of operations to increase the data quantity. To avoid data overfitting and get the most out of the models, we feed the improved data to Convolutional Neural Networks. [9] The issue of overfitting arises when a deep learning or machine learning model achieves good results on the training set but fails miserably on the validation and testing sets. Therefore, it is advised that the model not be overfitted during implementation. Expanding the CNN model's dataset allows us to better train Computer Vision (CV) and Deep Learning models, which in turn provide better, more organised results. Since this enables us to decrease the amount of disc space used for computations performed during runtime. Data augmentation techniques are deployed to improve the accuracy of our Deep Learning model is
Table 1: Data Augmentation Settings

<table>
<thead>
<tr>
<th>Data Augmentation Hyperparameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rescale</td>
<td>1./255</td>
</tr>
<tr>
<td>Shear range</td>
<td>0.3</td>
</tr>
<tr>
<td>Zoom range</td>
<td>0.3</td>
</tr>
<tr>
<td>Horizontal split</td>
<td>True</td>
</tr>
<tr>
<td>Colour mode</td>
<td>Grayscale</td>
</tr>
<tr>
<td>Validation split</td>
<td>0.6</td>
</tr>
</tbody>
</table>

1. Rescale – The term for this procedure is input normalisation. Scaling all of the images to the same range [0,1] will make the loss distribution more uniform. For pixels in the [0,255] range, setting rescale=1./255 to [0,1] will do the trick. Without scaling, the high pixel range photographs will mostly dictate how weights are updated. Even while a black-and-white Apple picture has more pixels than an all-black one, it doesn't necessarily make it better for training. Also, the model may process inputs faster since the neural network's coefficients are in the [0, 1] range instead of [0, 255]. This makes convergence more probable.

2. Shearing The term "shear" describes a distortion of an image along one or more of its axes, often used to create or fix perception angles. Usually, it's employed to make pictures seem better so that computers can simulate human perception of depth and other viewpoints. Potentially comparable test images do exist. A small number of shear-aligned training pictures. Do The shearing orientation may be determined only by analysing these test pictures. For improved understanding, use photographs from a designated training set.

3. Zoom range – The zoom augmentation method is used to zoom in on the picture. This method involves pixel-by-pixel randomization to expand or zoom in on a picture. This method makes advantage of the zoom range parameter of the Image Data Generator class.

4. Horizontal split – The image's dimensions stay the same when the pixels travel horizontally using a horizontal shift augmentation. Consequently, in order to keep the image's dimensions constant, certain pixels are clipped and new pixels are placed into a replacement zone. Horizontally dividing the image to True in this part allows us to create fresh data for the CNN Model.

5. Colour mode – The combination of black and white is what makes grayscale, another name for this colour mode. You may use it to make a copy of a picture with varying greyscale tones.

6. Validation split – After the data is shuffled, the model divides it into two sets: train and validation. With the train and validation datasets built up in the first epoch, the second epoch doesn't need to reshuffle and divide them; instead, it makes use of them. It follows that it is a confirmation.

Figure 4. Block diagram of the proposed method
Pre-processing Prior to being used for training or inference models, images must undergo processing. Some examples include changes in size, colour, and direction, among others. In order to conduct more accurate analyses, pre-processing raises the image's quality. We may improve some features that are important for our application and remove unwanted distortions by pre-processing. Depending on the use case, certain traits could change. Software has to pre-process an image before it can operate correctly and provide the desired results.

Deep learning models We have used the following deep learning models and compare their accuracies.

1) VGG16: To train the VGG-16 convolutional neural network model, researchers combed through the massive ImageNet dataset, which is used for picture classification and object recognition. Thirteen convolutional layers and three fully linked layers make up the VGG-16 architecture's total of sixteen layers. To train the model to categorise images, a thousand different classes were used. The very accurate picture classification capabilities of VGG-16 make it stand out among its many computer vision applications, such as object recognition, semantic segmentation, and transfer learning.

2) VGG19: Using a combination of convolutional and max-pooling layers, VGG19 analyses the input picture and extracts features. The final prediction is generated by the fully linked layers when they obtain these characteristics. By constantly using tiny convolutional filters (3x3) with a deep network architecture, the design effectively captures intricate picture features, which is its major benefit. A total of thirteen convolutional layers and three completely connected levels make up the design's nineteen layers. Among image classification tasks, VGG19 stands out for its exceptional accuracy and ability to process large image inputs.

3) RESNET-50: The design is unique in that it uses residual connections to allow deeper networks without running into the vanishing gradient issue. Fully connected, pooling, and convolutional layers are all part of ResNet-50's fifty layers. It has a stellar reputation for accurately classifying images and processing large amounts of incoming data.

4) RESNET-101: A deep convolutional neural network (CNN) design, ResNet-101 features more layers compared to ResNet-50. Similar to ResNet-50, ResNet-101 trains very deep networks using residual connections to avoid the problem of vanishing gradients. ResNet-101’s enhanced success on image classification tests might be attributed to its extra layers, which provide more representational capacity. Due to the increased computing power and resource requirements, the model's training time may increase as the number of layers increases.

As a result of its high mortality and morbidity rates, pneumonia remains a major healthcare concern, especially for at-risk groups including children and the elderly. In order to effectively treat and manage a patient, a prompt and correct diagnosis is essential. Delays in diagnosis might result in problems and higher healthcare expenses. Clinicians rely heavily on medical imaging, especially the study of chest X-rays (CXRs), to help diagnose pneumonia. The use of deep learning methods, especially CNNs, has transformed the field of medical image analysis in the last few years. Image classification, object identification, and segmentation are just a few of the many tasks where CNNs have shown to be very effective. Their usefulness in healthcare applications has been greatly enhanced by their capacity to automatically learn discriminative characteristics from raw data.

Convolutional neural networks (CNNs) have been very successful in medical image processing, and transfer learning—the process of applying what one learns in one area to another—has further added to this success. Effective feature extractors are pre-trained convolutional neural network (CNN) models that have been trained on massive natural image datasets like ImageNet. Improving performance with less labelled data and processing resources is possible by fine-tuning these models using medical imaging datasets, especially for applications like pneumonia diagnosis. Our paper presents a machine learning application that uses deep convolutional neural networks (CNNs) and transfer learning models to identify pneumonia in medical photos, particularly chest X-ray images. Our goal is to create a reliable technology that will help doctors diagnose pneumonia faster, which will improve healthcare for patients and doctors alike.

Conclusion

Using deep Convolutional Neural Networks (CNNs) and transfer learning models, we have developed a machine learning application that can identify pneumonia in medical photos. We were able to build
convolutional neural network (CNN) models using large-scale natural picture datasets, and then apply them to the problem of pneumonia diagnosis via transfer learning. Our model performance has been much enhanced, especially in cases when there is a lack of annotated medical data, thanks to transfer learning. We have created a strong and precise method for detecting pneumonia in medical photos by using deep convolutional neural networks (CNNs) and transfer learning. This tool will help improve patient outcomes and healthcare delivery. Both patients and doctors will reap the benefits of future studies that use machine learning to medical image processing, thanks to our work.

References


