UNIVERSITY OF SCIENCE AND TECHNOLOGY OF HANOI
INFORMATION COMMUNICATION TECHNOLOGY DEPARTMENT

Introduction To Deep Learning

Report: Pneumonia Classification from X-ray Images with Inception-V3

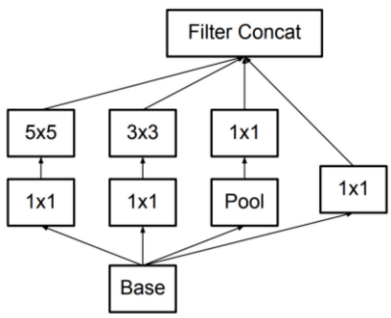
Group Information	**G38**
Le Xuan Loc	**22BI13256**
Nguyen Chi Nhan	**22BI13345**
Pham Duc Gia Hien	**22BI13155**
Le Duy Anh	**22BI13017**
Nguyen Duc Huy	**22BI13193**
Nguyen Quoc Khanh	**22BI13210**

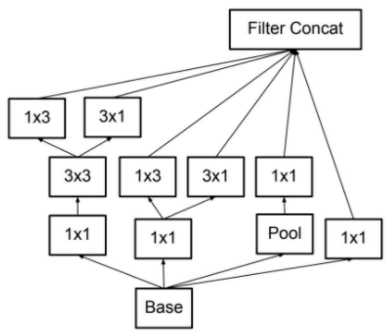
Abstract

Pneumonia is a serious infection of the respiratory system; it remains one of the leading causes of morbidity and mortality around the world, mostly among children and older adults. Early detection of the disease is of prime importance in order to reduce its burden. Traditional diagnostic approaches include chest X-rays and clinical examinations, which are time-consuming and may involve human error. The subsequent project deals with these challenges by designing an automated system for pneumonia detection in chest X-ray images using InceptionV3, a state-of-the-art CNN architecture.

1. Introduction

1.1 Background of InceptionV3





The Inception model was initially introduced in the article “Going deeper with convolutions” by Google researchers to enter the ImageNet Large Scale Visual Recognition Challenge 2014 (ILSVRC14). Researchers went on to try to improve the original v1 model and Inception-v2, Inception-v3, Inception-v4 … was born. The Inception-v3 model has some updates from the v1 in loss functions, optimizer, and batch normalization to make it more efficient.

1.2 InceptionV3 Overview

1.2.1 Description:

InceptionV3 is a deep convolutional neural network, also known as CNN, designed for image classification tasks. It is an improvement to the original Inception architecture that includes a number of changes to boost effectiveness and performance. The network can simultaneously gather data at several scales thanks to the architecture's utilization of Inception modules.

1.2.2 Advantage over other method

Inception V3 offers several advantages over other models, making it a highly effective choice for various tasks:

- Improved performance: Because of its capacity to learn multi-scale features, InceptionNet V3 achieves state-of-the-art performance on a number of image classification benchmarks, including the ImageNet dataset.
- Parameter efficiency: Compared to some other models, Inception v3 uses factorized convolutional algorithms to cut down on the number of parameters in a network. It utilizes smaller convolutions instead of bigger ones, like replacing one 5x5 filter (25 parameters) with two 3x3 filters (18 parameters).
- Flexibility: Because of its modular design, which makes expansions and alterations simple, it may be used for a wide range of tasks beyond image classification, including object identification and segmentation.
- Robustness to overfitting: The resilience of the model is increased by methods like global average pooling and auxiliary classifiers, especially when there is a shortage of training data.
- Lower computational cost: Because of its optimized computational efficiency, InceptionNet V3 can operate with low-resource devices and still achieve excellent accuracy.

1.3 Objective

The objective of this work is to create and evaluate an InceptionNet V3 model to detect pneumonia in chest X-rays. The objective is the accurate separation of healthy lungs from pneumonia-affected lungs using InceptionNet V3's powerful feature extraction capabilities. In order to aid medical professionals in accurately and promptly diagnosing pneumonia, this research attempts to reduce false positives and false negatives, enhance diagnostic accuracy, and provide a dependable, automated tool. Furthermore assessed will be the model's general efficacy, accuracy, and recall in clinical settings.

2. Proposed Methodology

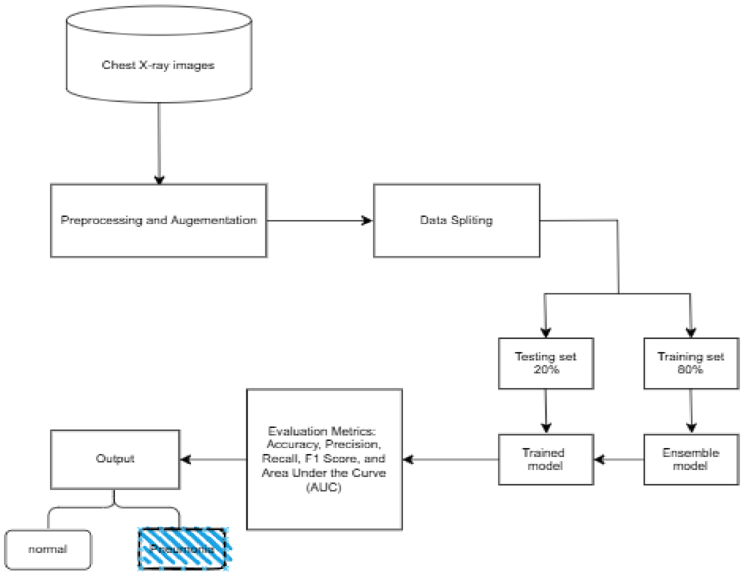
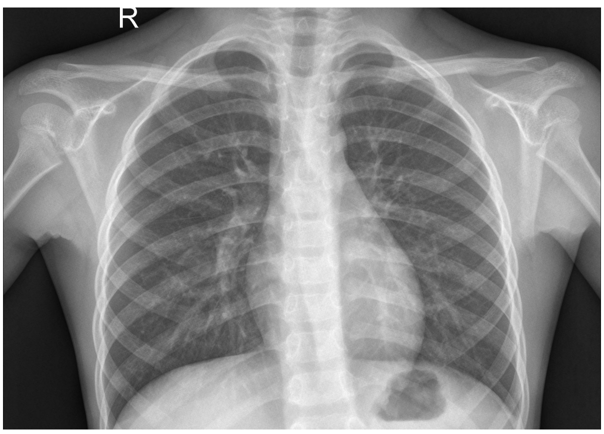


Figure. The workflow of the proposed methodology

2.1 Dataset description

This dataset contains Chest X-ray images of patients one to five years old taken from Guangzhou Women and Children's Medical Center. There are 5,563 images organized into 3 folders (train, test, val) and each one has two subfolders of two categories: normal patients and patients with pneumonia.

All radiographs are initially screened to control quality, removing low quality and unreadable scans. Then, two expert physicians graded the images' diagnoses before being cleared for training.



Normal

Pneumonia

2.2 Data Splitting

To train and assess the models in this work, the dataset is separated into training, testing, and validation sets. The testing set is utilized to assess the model's performance on unobserved data, whereas the training set is utilized to fit the model. A validation set is also utilized to adjust hyperparameters and avoid overfitting. 80% of the dataset is used for training, 20% is used for testing, and a tiny portion is used for validation. Table below provides information on the number of samples in each set for the two groups (normal and pneumonia).

Class	Training Set (80%)	Testing Set (20%)	Validation Set	Total
Normal	1,341	234	8	1,583
Pneumonia	3,875	390	8	4,273
Total	**5,216**	**624**	**16**	**5,856**

Table. Number of samples for train, test and validation split.

2..3 Inception

2.3.1 Freezing layers

In convolution neural networks (CNN), “layer freezing” is a technique used to control the way weights are updated. When a layer is “frozen”, the weights can no longer be modified. This technique is introduced to reduce the computational time for training while maintaining a fair amount on the accuracy side. In addition, freezing a layer can accelerate neural network training by progressively freezing hidden layers. In this particular model, it freezes all layers but the final one, which means we only need to update the weights of the last layer.

2.3.2 Modification of the final layer

In the InceptionNet v3 model, the final fully connected layer takes the output of prior layers, then converts them into a set of class probabilities for classification. To train this dataset, we modified the final layer of the pre-trained model - which was trained on a specific number of classes - with a new layer that fits this dataset. In this case, the number of classes is 2.

2.3.3 Training parameters

- CrossEntropyLoss (also known as log loss): a formula used to measure the performance of a classification model. It ranges from 0 to 1 and the lower the score is, the better the performance of our model. For classification tasks, cross entropy measures the difference between the predicted probability and true probability. There are 2 variants of cross-entropy loss: Binary Cross-Entropy Loss and Multiclass Cross-Entropy Loss.
- Adam Optimizer (Adaptive Moment Estimation): an algorithm widely used in machine learning and deep learning, it can be considered a combination of Momentum and RMSprop, which are two other optimizers. Some of its important features include: Adaptive Learning Rates, Momentum, and Bias Correction. It works by customizing each parameter's learning rate based on its gradient history; this adjustment helps the neural network learn more efficiently as a whole. There are a number of benefits from Adam Optimization such as: faster convergence, easy to implement and little memory requirements.
- An epoch is defined as the total number of iterations of all the training data for a machine learning or deep learning model. The number of epochs is considered a hyperparameter, it represents the number of times the whole dataset has run through the learning model. In this model, we take 10 epochs, which is a relatively smaller number of epochs compared to other models. From these epochs, we are able to plot a learning curve to see insights as to how fit the model is.

2.4 Training Setup

In this study, we leveraged transfer learning by utilizing a pre-trained deep learning model for the pneumonia detection task. This approach allows us to avoid training a deep neural network from scratch, which is often computationally expensive and susceptible to overfitting due to the limited size of the training dataset.

2.4.1 Pre-training

Using a pre-trained convolutional neural network (CNN) model that has been trained on a large dataset allows the model to recognize hierarchical features in images. These learned features, such as edges and shapes, can be effectively transferred to our specific task of detecting pneumonia in chest X-rays. The fine-tuning process primarily adjusts the weights of higher layers while retaining the knowledge of lower-level features.

2.4.2 Key Parameters

- Optimizer: The optimizer used for training is Adam, which adjusts the learning rates based on the first and second moments of the gradients, providing fast convergence.
- Learning Rate: A learning rate of 0.001 was initially set and reduced using a learning rate scheduler if no improvement in the loss function was observed.
- Epochs: The model was trained for 50 epochs to ensure it could effectively learn patterns while avoiding overfitting.
- Batch Size: A batch size of 32 was chosen to provide a good balance between memory consumption and training speed.

2.4.3 Loss Function and Evaluation Metrics

Loss Function is a categorical cross-entropy loss function, which is frequently used for multi-class classification issues, is utilized to optimize the model. This loss function effectively handles scenarios where the model is required to categorize the data into numerous categories, making it an appropriate choice for the pneumonia dataset.

Several evaluation metrics were used to evaluate the model's performance. Four alternative outputs are identified by the model: false positive (FP), false negative (TN), true positive (TP), and false negative (FN). Pneumonia is represented by TPs, or correctly predicted positive instances, whereas normal cases, or appropriately identified negative instances, are represented by TNs. Conversely, FPs and FNs represent inaccurate categorizations in which negative instances are mistakenly classified as positive and vice versa.

The following metrics were used to evaluate the model:

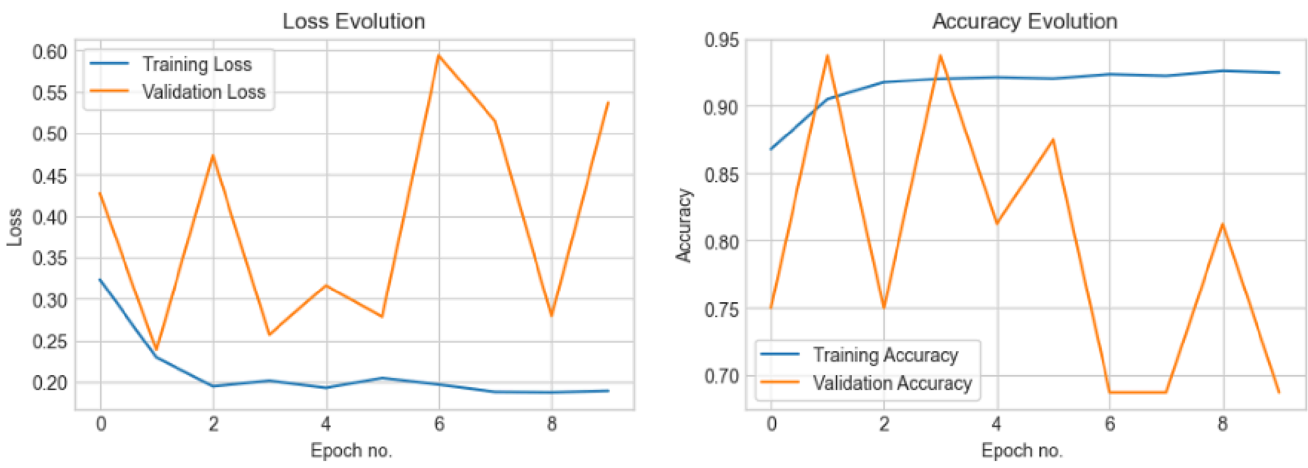
- Accuracy: The ratio of correctly predicted cases (both pneumonia and normal) to the total number of predictions, providing an overall assessment of the model's performance.

- Precision: The ratio of true positives to the sum of true and false positives. It indicates the quality of the model's positive predictions by quantifying how many of the predicted pneumonia cases were actual pneumonia cases.
- Recall: The ratio of true positives to the sum of true positives and false negatives. This metric reflects the model's ability to correctly identify all true pneumonia cases.
- F1-Score: A harmonic mean of precision and recall, offering a single metric that balances both precision and recall values, particularly useful when dealing with imbalanced datasets.
- Cohen's Kappa: This statistic measures the agreement between predicted and actual classifications, accounting for the possibility of random chance.
- AUC (Area Under the ROC Curve): A performance measurement that captures the model's ability to differentiate between classes across different threshold settings.

By leveraging these metrics, the performance and reliability of the model on the pneumonia dataset were thoroughly evaluated, with the pre-trained models and these key parameters enhancing both training efficiency and model accuracy.

3. Result and Discussion

Training and Validation Results:

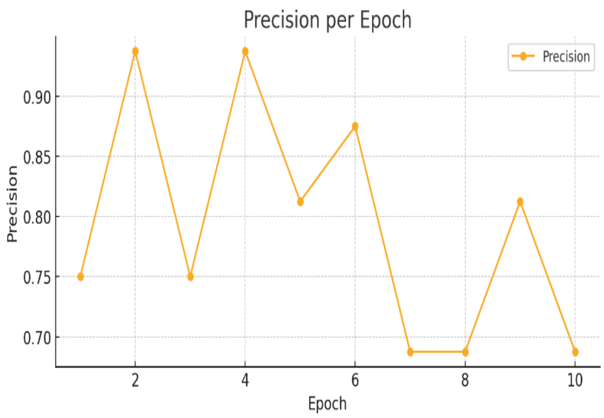


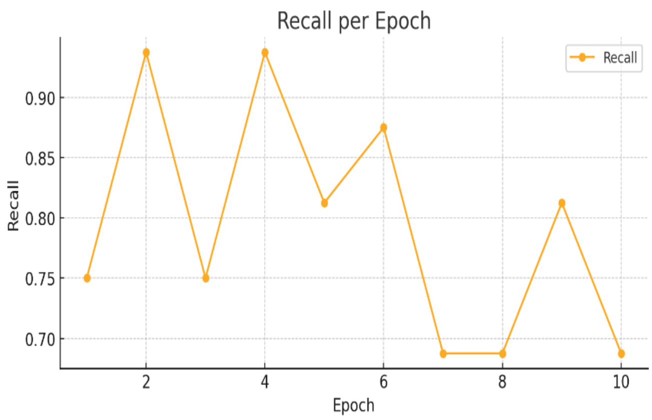
From the results illustrated in the graph:

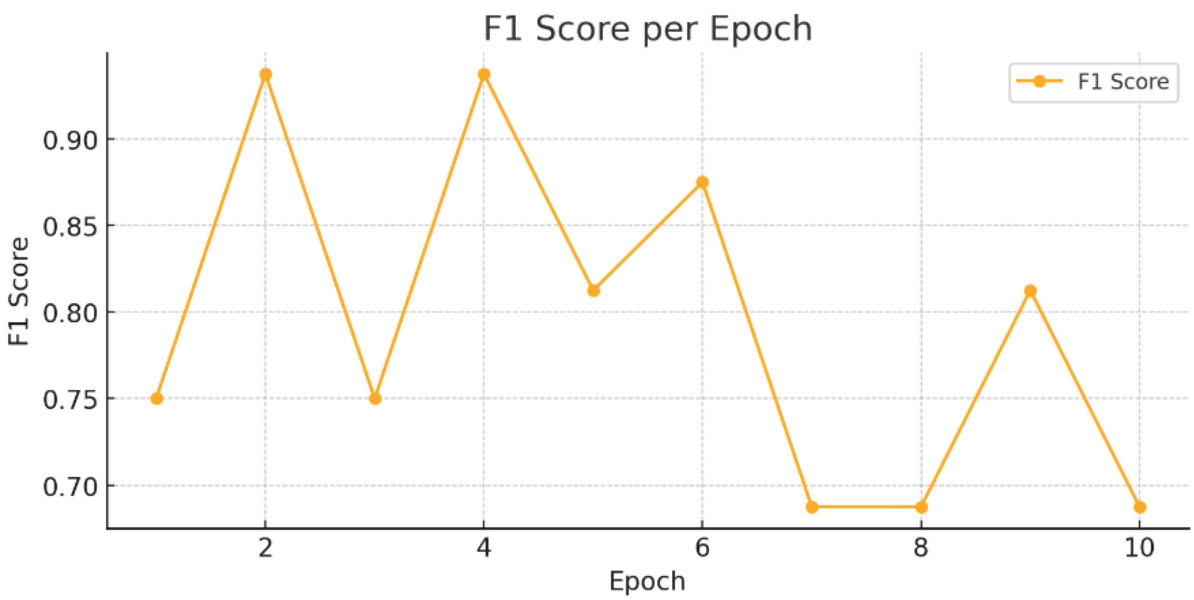
- Training Loss consistently decreased over the epochs, indicating the model was learning well on the training data.

- Validation Loss fluctuated, showing overfitting trends. Although it improved in some epochs, it worsened in others (especially in epochs 3, 7, 8, and 10), suggesting that the model may not generalize well on unseen data.
- Training Accuracy steadily improved, reaching 92.48% by epoch 10.
- Validation Accuracy fluctuated more, peaking at 93.75% in some epochs but also dropping to 68.75% in others. These results illustrate that while the model has a good performance on training data, it struggles to maintain consistent efficiency on the validation set.

Performance Metrics:







- The precision, recall, and F1 scores followed a similar pattern to accuracy, with significant fluctuations between epochs. While these scores reached a high of 93.75% during some epochs, they also dropped to as low as 68.75% in others.

Analysis of Misclassifications:

Discussion of errors or misclassified examples and potential reasons.

- Overfitting: Validation accuracy fluctuates, suggesting the model may be memorizing training data rather than generalizing.
- Complex/Noisy Data: Possible errors due to ambiguous features or noise in the data.

Insights on how the model could be improved.

- With complex/noisy data, we could use smoothing techniques to reduce noise in the data by averaging for filtering values, making the underlying more apparent, and outlier detection methods to identify data points that significantly deviate from the majority of the dataset.
- In the future, the overfitting problem may be solved by increasing the batch size.
- Increasing the training data amount by applying transformations.
- Use k-fold cross validation to ensure the model generalizes well on different subsets of the dataset.

4. Conclusions

Our findings show that InceptionV3 is a valuable tool for identifying pneumonia, underscoring its potential to improve diagnosis accuracy in healthcare. It is an excellent fit for medical applications and may help professionals diagnose patients effectively thanks to its superior picture recognition capabilities.

By simplifying diagnosis, the method may speed up patient care in practical situations, particularly in areas with limited medical resources. Subsequent research ought to concentrate on utilizing increasingly intricate datasets to enhance resilience and exploring advanced tactics for improved outcomes. Expanding the model's identification to include more medical conditions could boost its impact and encourage the development of AI-powered healthcare innovations.

5. References

[A Simple Guide to the Versions of the Inception Network | by Bharath Raj | Towards Data Science]

[What Does Freezing A Layer Mean And How Does It Help In Fine Tuning Neural Networks (analyticsindiamag.com)]
[What Is Cross-Entropy Loss Function? - GeeksforGeeks]
[Going Deeper with Convolution | by Wei Liu and 8 others]