Computer Systems Science and Engineering

Tech Science Press

DOI: 10.32604/csse.2023.042078

ARTICLE

Optical Based Gradient-Weighted Class Activation Mapping and Transfer Learning Integrated Pneumonia Prediction Model

Chia-Wei Jan[1], Yu-Jhih Chiu[1], Kuan-Lin Chen[2], Ting-Chun Yao[3] and Ping-Huan Kuo[1,4,*]

[1]Department of Mechanical Engineering, National Chung Cheng University, Chiayi, 62102, Taiwan

[2]Department of Intelligent Robotics, National Pingtung University, Pingtung, 900392, Taiwan

[3]School of Medicine, College of Medicine, Taipei Medical University, Taipei City, Taiwan

[4]Advanced Institute of Manufacturing with High-Tech Innovations (AIM-HI), National Chung Cheng University, Chiayi, 62102, Taiwan

*Corresponding Author: Ping-Huan Kuo. Email: phkuo@ccu.edu.tw

Received: 17 May 2023 Accepted: 24 July 2023 Published: 09 November 2023

ABSTRACT

Pneumonia is a common lung disease that is more prone to affect the elderly and those with weaker respiratory systems. However, hospital medical resources are limited, and sometimes the workload of physicians is too high, which can affect their judgment. Therefore, a good medical assistance system is of great significance for improving the quality of medical care. This study proposed an integrated system by combining transfer learning and gradient-weighted class activation mapping (Grad-CAM). Pneumonia is a common lung disease that is generally diagnosed using X-rays. However, in areas with limited medical resources, a shortage of medical personnel may result in delayed diagnosis and treatment during the critical period. Additionally, overworked physicians may make diagnostic errors. Therefore, having an X-ray pneumonia diagnosis assistance system is a significant tool for improving the quality of medical care. The result indicates that the best results were obtained by a ResNet50 pretrained model combined with a fully connected classification layer. A retraining procedure was designed to improve accuracy by using gradient-weighted class activation mapping (Grad-CAM), which detects the misclassified images and adds weights to them. In the evaluation tests, the final combined model is named Grad-CAM Based Pneumonia Network (GCPNet) out performed its counterparts in terms of accuracy, precision, and F1 score and reached 97.2% accuracy. An integrated system is proposed to increase model performance where Grad-CAM and transfer learning are combined. Grad-CAM is used to generate the heatmap, which shows the region that the model is focusing on. The outcomes of this research can aid in diagnosing pneumonia symptoms, as the model can accurately classify chest X-ray images, and the heatmap can assist doctors in observing the crucial areas.

KEYWORDS

Data augmentation; deep neural network; image classification; medical imaging; transfer learning; Grad-CAM

1 Introduction

Artificial intelligence (AI) is gradually becoming widely used in medical diagnosis. A study developed an AI-integrated app that provides reminders about the user's dental conditions and reduces the time spent for clinical examination [1]. Because the traditional scanning method is expensive, another study proposed a new model for Glaucoma detection. The model had a lower negative rate and was as accurate relative to manual scanning and is suitable for mass scanning [2]. AI-based image enhancement techniques are also being implemented; Because these technologies can enhance the features of images and promote the model's ability to analyze images. For example, reference [3] used parameters such as contrast and brightness as input and enhances image features through neural networks.

Even before COVID-19, pneumonia was a common infection. There are 150.7 million new cases of COVID-19 in developing countries yearly [4]. Hence, developing a integrated pneumonia diagnosis system is important. Researchers have proposed a variety of antipandemic measures. In [5], a three-dimensional lung segmentation approach that improves the efficiency of image processing was proposed. Deep neural networks and especially deep convolution neural networks (CNN) applied to medical imaging are becoming increasingly popular. Researchers have also proposed a brain tumor segmentation method using CNN, which uses small kernels [6]. In addition, this reference [7] investigated the performance of various optimizers of the CNN model. These include Stochastic Gradient Descent (SGD), Adaptive Gradient Descent (Adagrad), Adadelta, Root Mean Square Propagation (RMSprop), and Adam. The study showed that the model that using Adam optimizer had an accuracy rate of over 99%. SqueezeNet-Guided ELM (SNELM) [8] was used to diagnose COVID-19 in this study. This method also used data augmentation to expand the dataset and used SqueezeNet (SN) to generate SN features. The study showed that the model had fast learning speed and good generalization performance, making it feasible for diagnosing COVID-19.

Researchers have found it difficult to improve prediction accuracy due to the lack of available medical imaging data, because medical images are more difficult to obtain than other types of images. Furthermore, medical imaging data must be labeled by many doctors to be useful for machine learning, which entails a high cost in high time and resources.

Data augmentation techniques can be used to solve the problem of data scarcity. Reference [9] used CNN architecture and data augmentation for sound classification. They also evaluated different augmentation techniques against each other. Data augmentation is also used to expand the images of brain tumors [10]. The article mentions that the study completed data augmentation. Machine learning becomes more effective when trained on large datasets. However, obtaining large datasets can be challenging. Therefore, performing data augmentation is very helpful for training. From a practical perspective, it should be noted that not all normal X-rays and pneumonia X-rays exhibit variations in brightness due to unique characteristics of each X-ray machine. Therefore, data augmentation is necessary to address this issue.

The model may overfit when there is insufficient training data available in the target domain. Fortunately, transfer learning solves this problem [11]. Much research on transfer learning has been conducted in the past decade, especially in visual categorization such as image classification [12]. Furthermore, transfer learning can be implemented in a variety of ways [13], all of which are based on instances, features, parameters, or relations. In [14], the researchers provided a comprehensive summary of more than 40 typical methods of transfer learning. Reference [15] proposed a transfer learning framework for pneumonia detection, which compares the different combinations of pre-trained models and classifiers. Another study applied the model compression technique after transfer learning, resulting in decreased hardware requirements [16].

The deep neural network is akin to a black box. Therefore, many researchers have focused on visual explanations. In CNN models, fully connected layers are replaced by global average pooling (GAP), which increases interpretability and prevents overfitting [17]. A procedure to generate class activation maps (CAM) is designed based on GAP, wherein the weighted sum of feature maps is calculated [18]. This technique highlights the principal area detected by CNN. However, CAM cannot be used when GAP is not in the model, such as the Visual Geometry Group (VGG). This task is solved by using Grad-CAM [19]. Grad-CAM is a generalization of CAM, which can be used for a wide range of CNNs, especially for fully connected layers.

Grad-CAM is quite a practical method. It can analyze the region that the CNN model focuses on. If the CNN model does not focus on the lungs, it will be highly possible to cause wrong judgment. It is hoped that in this research, the center of the image, which corresponds to the lungs, can be focused on by CNN. If the region CNN focuses on is not at the center of the chest, we will train the image again to improve the accuracy of CNN's prediction. Grad-CAM can analyze the areas that the current CNN model focuses on. If the area that the current CNN focuses on for a certain image is not in the center of the image, it may indicate that there is a possibility of incorrect judgment for this image. During the training process, it is not possible to control CNN to forcibly focus on the center of the image. Therefore, if the area that CNN focuses on is in the wrong position, the image will be retrained. This process will be repeated until the training error is less than a certain standard.

In addition, Simple Vision Transformer (SimViT) [20] can also integrate spatial structure and local information into visual transformers. At the same time, SimViT extracts multi-scale hierarchical features from different layers for dense prediction tasks. Research shows that this method is suitable for various image processing tasks.

In [21], researchers proposed an automated methodology of lung diagnosis from chest X-ray images. The model was designed to identify and categorize typical viral pneumonia and COVID-19. Grad-CAM was used to detect where the model focuses on. In [22], because COVID-19 data were inadequate in 2020, transfer learning based on related chest X-ray data sets was used to detect COVID-19. At the end of transfer learning, Grad-CAM was also used for visualization. In [23], U-Net was used for the classification model. The segmentation task is executed if the model classifies the image as one of COVID-19 or pneumonia. The area is segmented based on the output of Grad-CAM.

Researchers in [24] have developed a decision-support and segmentation system for COVID-19, which uses EfficientDet and EfficientNet for image classification and segmentation. Moreover, the system can reject unrelated images using header analysis and the classifier. Another application is to use machine learning and optimization algorithms to predict insurance premiums before and after COVID-19 [25]. The study used Fast Approximate Nearest Neighbors (FLANN) and genetic algorithms and analyzed asymmetric data using five years of data to predict future costs. The study is superior to other methods.

A study has been conducted on generating hash values using convolutional stacked denoising autoencoder (CSDAE) [26]. This research can shorten the hash program while maintaining machine efficiency. Currently, many image processing-related technologies have been developed [27,28].

This paper uses the pretrained models VGG16, and ResNet50 for transfer learning. The model can classify normal and pneumonia chest X-ray images. In addition, using Grad-CAM to design a retraining procedure in this study improves the model's accuracy. Because if the area of interest of a certain data is not in the center of the picture, it means that the recognition performance of the picture may not be good, so the model will increase its weight in the next training process.

Machine learning models were integrated into a new system for pneumonia diagnosis in this study, which improved the accuracy of diagnosis. The data's balance between different classes was effectively achieved through data augmentation. This approach can avoid incorrect prediction results caused by data imbalance during the training process, in which the model predicts the class with more data as much as possible. A smart pneumonia diagnosis system is proposed in this paper. This system can analyze X-ray images and provide diagnostic suggestions for medical personnel. In addition to providing medical personnel with a faster way to identify the main causes of patients' illnesses, this approach can also improve diagnosis efficiency. Even when hospitals are too busy, medical personnel may make incorrect judgments due to poor mental conditions. If the method proposed in this paper can be used, it can effectively reduce the occurrence of diagnostic errors, thereby improving the quality of medical care.

An integrated pneumonia diagnosis system was proposed in this paper, which integrates multiple models and algorithms, and significantly improves the accuracy of pneumonia diagnosis. The system can quickly diagnose whether there is inflammation in the lungs and provide medical personnel with reference. The system will help alleviate the shortage of medical resources, improve diagnostic efficiency, and thus enhance medical quality. In the future, it can be practically applied in clinical medicine.

Although many existing methods were used in this study, integrating these technologies through existing methods can improve recognition efficiency. By integrating existing methods, the advantages of each approach can be combined to obtain more accurate judgment results. A more rigorous method is required for medical conditions to make correct judgments, and the abilities of different excellent methods can be combined by the method used in this study, achieving better performance than the original model. After being trained by machine learning, it is usually directly applied to the corresponding tasks. However, this approach may not be able to check whether known knowledge has been correctly absorbed by the model. To solve this problem, Grad-CAM is used to confirm whether the correct area is focused on by the CNN model. If not, the training for that data will be strengthened to enhance the predictive model's performance. Data augmentation can improve the generalization ability of the model, allowing the model to make judgments when facing data from other datasets. When applying cross-validation, the dataset is initially divided into ten parts. Data augmentation is used for training data rather than testing data in these ten parts, thus avoiding the data leakage problem. In summary, this study integrated existing technologies and fine-tuned their models for training. The GCPNet has excellent performance in judging pneumonia. This is a significant contribution to clinical research.

The rest of the paper is organized as follows. Section 2 presents the model architecture to classify X-ray images using CNN and transfer learning. Section 3 presents the performance evaluation results. Section 4 is the discussion of the proposed method. Finally, Section 5 concludes the paper.

2 Model Architecture

This study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Human Research Ethics Committee, National Chung Cheng University (Application number: CCUREC111090601). The bias is required to prevent relying on one particular system of training and testing data sets. For example, if the testing data is the same as the training data, the model cannot prove that it can predict well for other data even if the accuracy is 100%. Therefore, a stricter method is required to evaluate the model.

Cross-validation splits the original data set into training and testing data and calculates the average result of different partitions. Here 10-fold cross-validation is used to evaluate the performance. 10-fold cross-validation means the original data set is partitioned into ten sets. To train the model, 9-fold cross-validation is used while 1-fold cross-validation is used for testing. The results of ten iterations are averaged. Fig. 1 illustrates the 10-fold cross-validation.

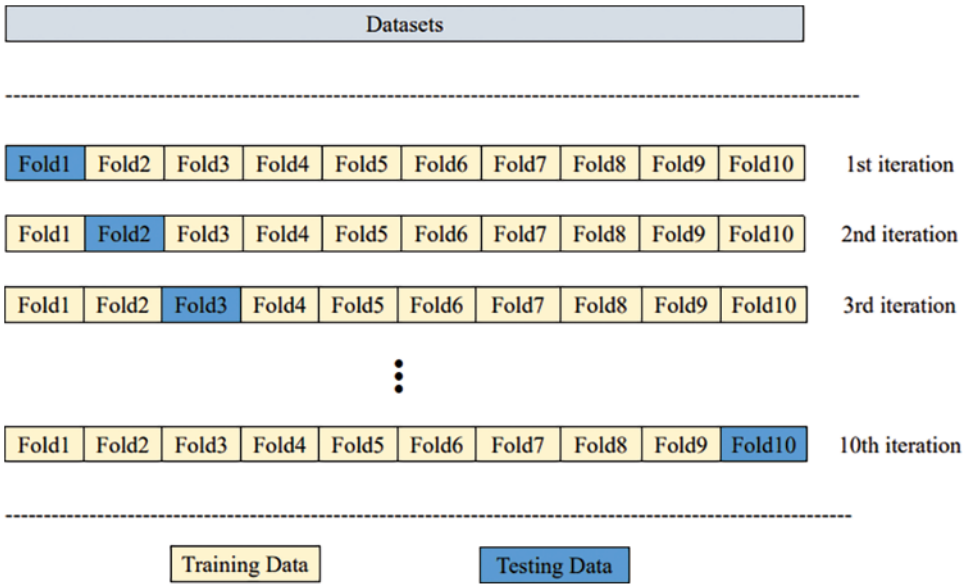


Figure 1: Ten-fold cross-validation

2.1 Transfer Learning

Collecting and labeling data is both challenging and time-consuming. The model is designed to be generalizable and not a one-to-one model. Transfer learning is implemented to help ensure this. Transfer learning focuses on storing features gained through one problem and applying them to another different but similar problem. Pretrained models such as VGG16 and ResNet50 can be used instead of building a neural network model from scratch to solve the problem. These models are usually pretrained using large-scale data sets. The model is composed of an extractor and classifier (Fig. 2).

In this paper, transfer learning is implemented with VGG16 and ResNet50, which pretrains the model with the ImageNet data set, which has over 14,000,000 images and 21,000 categories.

The structure of VGG16 is shown in Fig. 3. It has five blocks containing convolution layers and max pooling layers. Finally, it uses three fully connected layers to connect with the output layer.

In this paper, a VGG16 model, which has been pretrained with the ImageNet data set, is loaded. However, the entire model cannot be used because it has 1,000 outputs and only two outputs exist in this data set. Thus, the top of the model, which consisted of fully connected layers, was removed and a new output layer was added.

When pretrained models are used to implement transfer learning, some specific layers are frozen from the training process; thus, the weights are initially maintained at their ImageNet values and do

not change with backpropagation. Therefore, the frozen layers maintain the features of ImageNet, and the trainable layers can be improved with a new dataset.

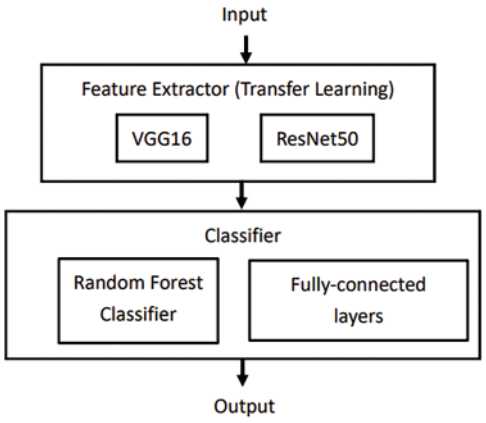


Figure 2: Structure of the model composed of extractor and classifier

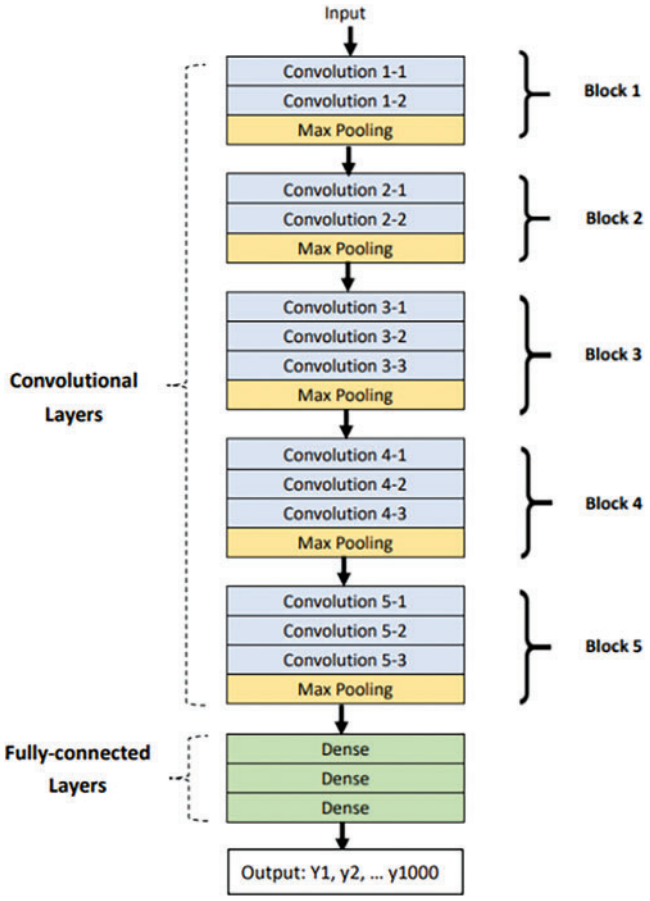


Figure 3: Structure of VGG16

Transfer learning involves two processes: feature extraction and fine-tuning. In the first stage, feature extraction is used to extract useful features from pretrained models and train new datasets only on fully connected layers. The model does not have to be retrained entirely because the base convolution layers already have some useful features. In the second stage, fine-tuning is used to unfreeze a few layers of a pretrained model for the last base convolution layers and the new classifier to be trained. This step allows us to tune some features in the base model. Feature extraction and fine-tuning is shown in Fig. 4.

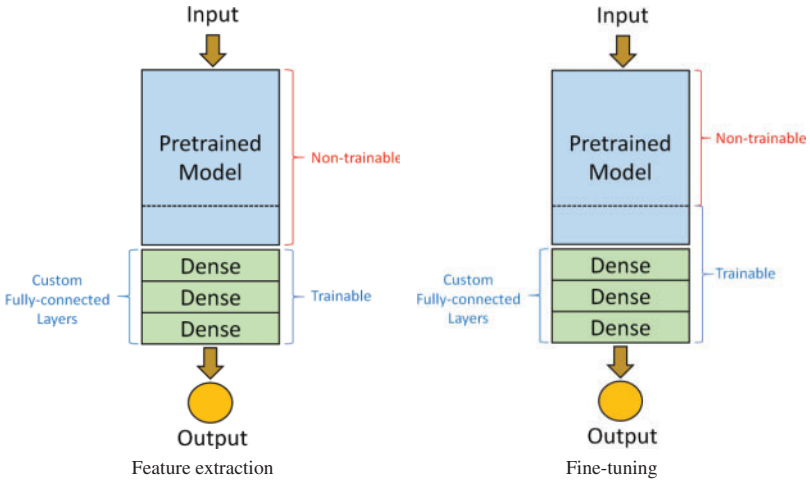


Figure 4: Two-stage training process

ResNet50 is considered to be the next generation iteration of VGG16. Degradation may happen when more layers are added to increase performance. For transfer learning, most of the network layers of ResNet-50 were utilized in this study. At the beginning of training, only the weights of the last few layers can be updated. Fine-tuning is to allow more weights so that they can be updated during the training process. The two-stage approach can increase training efficiency and achieve better recognition performance. ResNet50 solves this problem and adds more layers without degradation. With the skip connection, the network can learn identity mappings more easily. Therefore, the model does not degrade with residual block. The skip connection is shown in Fig. 5.

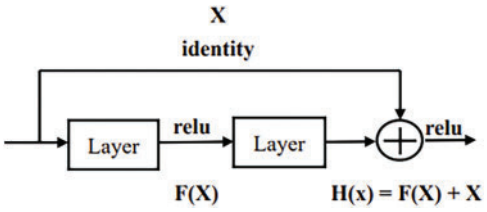


Figure 5: Skip connection

2.2 Grad-CAM

Grad-CAM is a method for understanding the basis of classification in convolutional neural networks (CNNs). Grad-CAM calculates the weight of each feature map in the last convolutional layer for the image category, then calculates the weighted sum of each feature map, and finally maps the weighted sum of the feature maps to the original image. Unlike CAM, which requires GAP in

the model, Grad-CAM can be used for a wide range of CNN models. Therefore, Grad-CAM is more flexible than CAM and can be used in most neural networks, such as fully connected layers, RNN, and long short-term memory, without the need to revise the model. Crucially, Grad-CAM can calculate the weighting through back propagation, as written in (1).

$$\omega_k^c = \sum_i \sum_j \frac{\partial Y^c}{\partial A_{ij}^c} \tag{1}$$

where Y^c is the classification score for class c, A is the feature map, ω_k^c is the class feature weights, and k is the number of different feature maps. An ReLU is applied to the linear combination in (2), because only the positive influence is of interest. α_k^c is the neuron importance weights.

$$L_{Grad-CAM}^c = ReLU\left(\sum_k \alpha_k^c A^k\right) \tag{2}$$

The output of Grad-CAM is a heatmap. In the heatmap, the region with the highest value is the region most focused on by the model.

The heatmap is divided into 25 blocks (Fig. 6). If the highest value is not located in the center, the model did not focus on the right position. Consequently, the image is given additional weight. The focus of the study is indeed on the center of the picture. However, there are many organs in the chest, and the heart and lungs are also located in the chest. However, due to the different body shapes of each person and the slight differences in the angle of X-ray photography, so it cannot completely exclude the location of the heart and focus only on the lungs. Some diseases also have correlate with the features exhibited by the heart and lungs. Therefore, it is less feasible to deliberately exclude the heart part. In summary, it is reasonable and practical for the model to focus on the center point of X-ray.

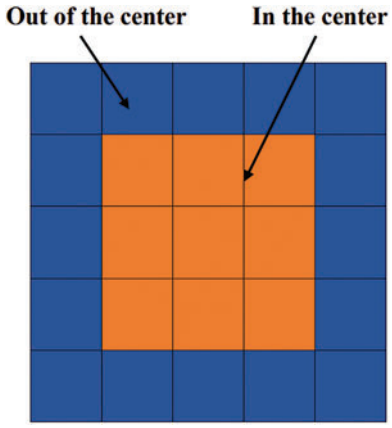


Figure 6: Division of the heatmap into 25 blocks

Superimposed visualization can be created by superimposing the heatmap onto the original image. Heatmaps are shown in Fig. 7. Superimposed images are shown in Fig. 8.

The proposed method is shown in Fig. 9. ResNet50 was used in this study with the Adam optimizer, a learning rate of 0.001, decay steps of 755, and decay rate of 0.9. The loss function used was binary cross entropy. The output was then passed through Grad-CAM to generate a heatmap. The generated heatmap will be evaluated and the image weights updated in subsequent training. In the first step, data augmentation is applied to balance the data. In step 2, the data is preprocessed by

normalizing the features between 0 to 1 and resizing the images to (160, 160). In step 3, the model composed of the ResNet50 pretrained model and fully connected layers starts the training process. In step 4, Grad-CAM is used to generate the heatmap of the training data. The heatmap is evaluated in step 5. If the highest value of the heatmap is off center, the weight of the image is increased. In step 6, the model is trained again. Finally, the model is evaluated by the confusion matrix.

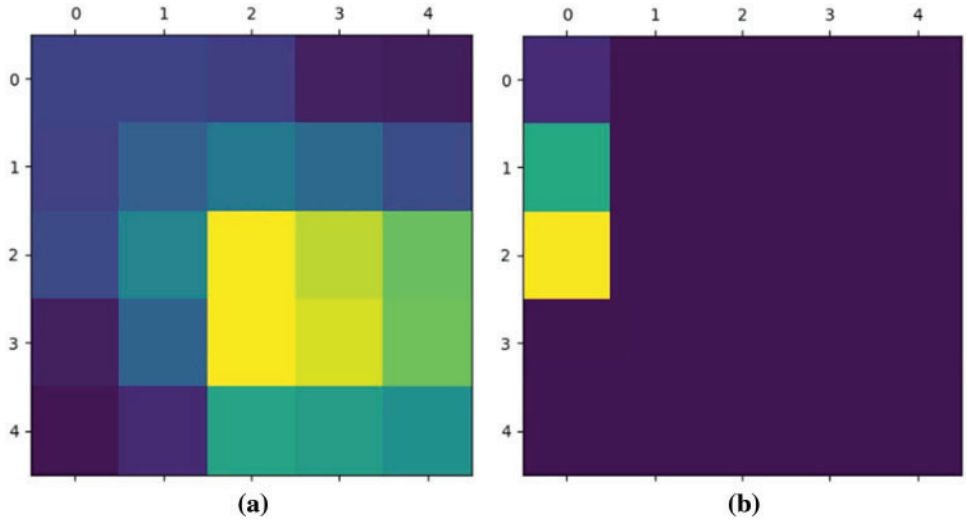


Figure 7: The highest value of heatmap is (a) in the center, (b) out of the center

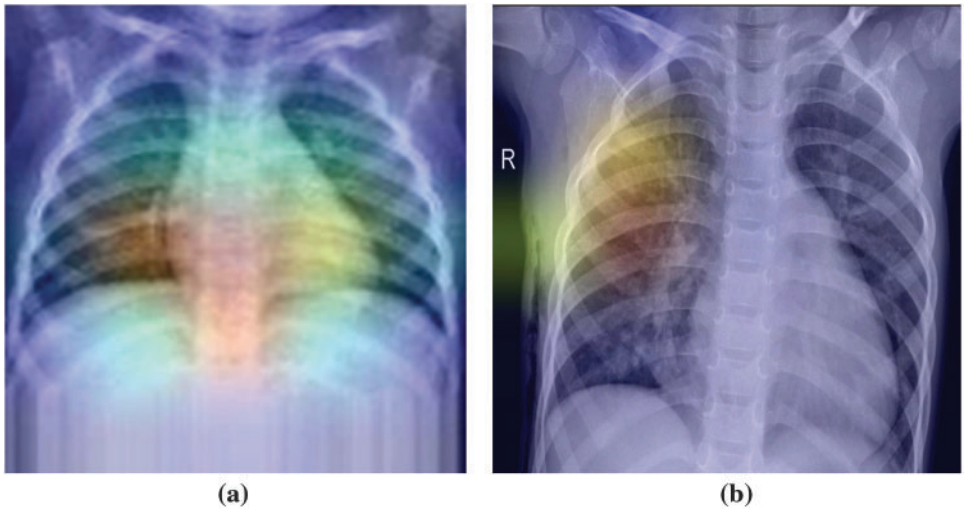


Figure 8: The highest value of superimposed images is (a) in the center, (b) out of the center

Although many existing models were used in the method proposed in this study, when these technologies were integrated into a system, they could effectively analyze pneumonia data. Through data augmentation, the proposed model can also perform well when facing data from other datasets in the future.

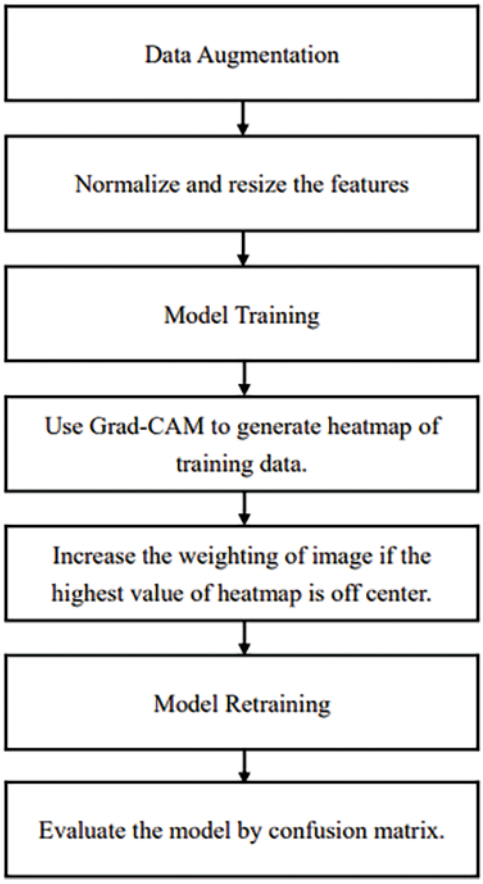


Figure 9: The proposed method

3 Performance Evaluation

3.1 Data Preprocessing

This dataset comprised chest X-ray images of lungs obtained from [29,30]. These images were selected from patients from Guangzhou Medical Center. The dataset is from an open dataset on the Internet. Generally, to obtain hospital images, they must be reviewed and approved by the patient before use. The image is in a standard image format (jpeg) and can be converted to a tensor format for use. The images were categorized into "normal" and "pneumonia" images (Fig. 10).

The data set had 1,341 and 3,872 normal and pneumonia images, respectively (at a ratio of 1:3). If the model is trained on imbalanced data, the model is unreliable. This is because the model tends to output guesses for the imbalanced category (with more data points) during training, thus performing poorly on a real-world data set. In addition to this data imbalance problem, the total amount of images was 5,213. The model may tend to learn the category with the higher frequency rather than how to discriminate between categories when making judgments.

Data augmentation is used to solve the problems of data imbalance and data insufficiency and is a useful technique for increase data diversity through image transformations, such as shifting, flipping, rotating, zooming, and brightening.

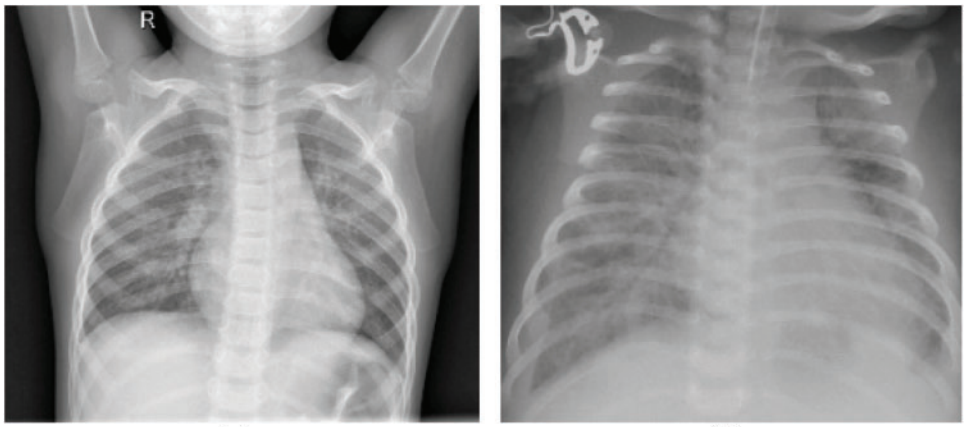


Figure 10: X-ray image of (a) normal person, (b) pneumonia person

To prevent unrealistic images, the shift, brighten, rotate, and zoom functions were used instead of flipping real X-ray images. The angle of rotation was limited to 15°. The images after rotating, horizontal shifting, vertical shifting, and brightening or darkening are shown in Figs. 11–14. The data augmentation setting is described in Table 1.

After data augmentation, the data set had 6,000 normal images and 6,000 pneumonia images (Fig. 15).

Before the model is trained, the images should all have the same pixel count. Data resizing is used to minimize all images to (160, 160). To enable the neural network model to read the images, the image type was converted into NumPy array instead of flipping real X-ray images. The images of normal and pneumonia patients after resizing and transforming them into a NumPy array are shown in Fig. 16. Furthermore, normalization should be done before model training. Since the gray level of the images is between 0 to 255, the NumPy arrays are divided by 255.

Figure 11: Horizontal shifting

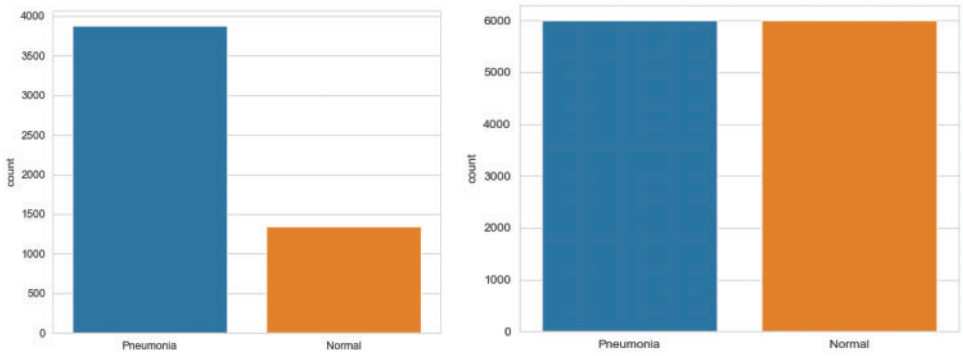
Figure 12: Vertical shifting

Figure 13: Rotation

Figure 14: Brightening and darkening

Table 1: Data augmentation setting

Data augmentation techniques	Value
Width shift range	0.15
Height shift range	0.15
Brightness range	0.2
Rotation range	15



Before data augmentation After data augmentation

Figure 15: Number of images

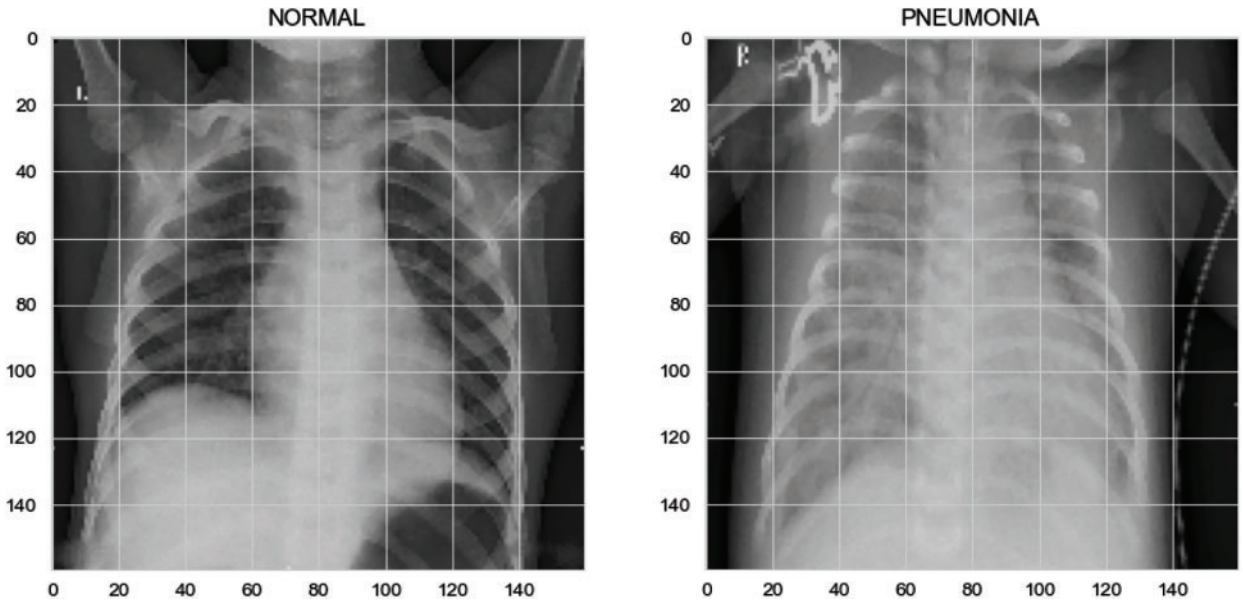


Figure 16: Image resizing

The model's input is the X-ray images after data preprocessing, and the output is the label of pneumonia or normal. Data encoding is used to convert categorical data into numerical values that

the model can read. The category "Pneumonia" is labeled as "0" and the category "Normal" is labeled as "1."

3.2 Loss and Accuracy

The results on the loss and accuracy of various models in evaluation tests are presented in this section. The epochs were set to 25 for each model. Moreover, the fine-tuning process starts at the fifth epoch onward. The loss and accuracy of the models are illustrated in Figs. 17 and 18.

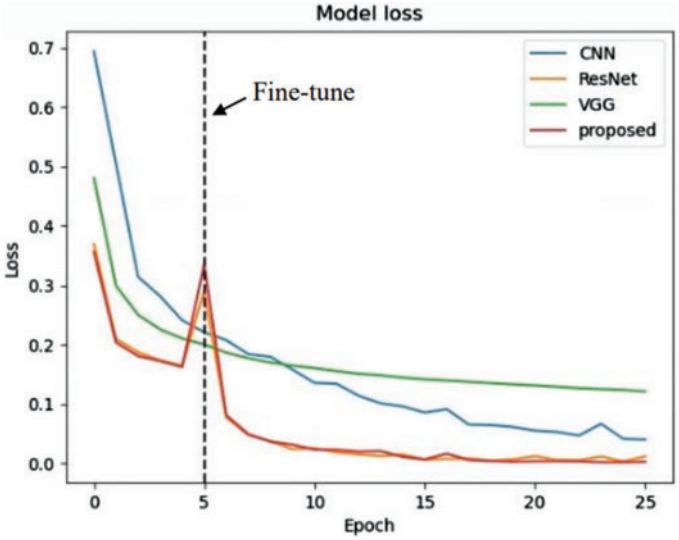


Figure 17: Learning curve (loss)

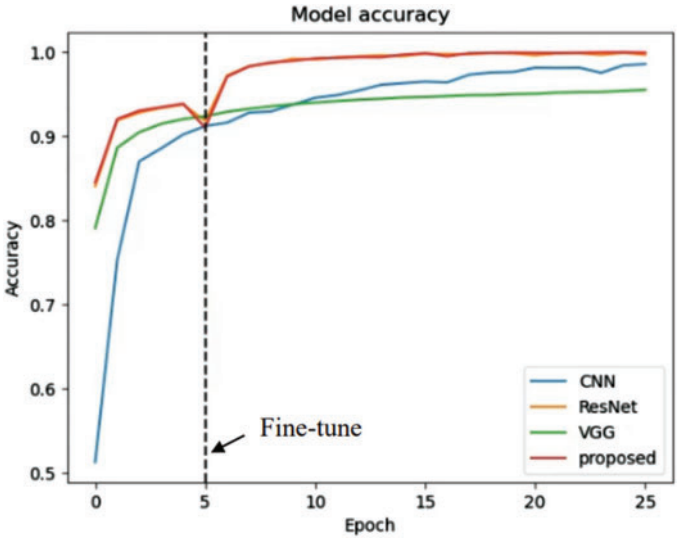


Figure 18: Learning curve (accuracy)

In the evaluation tests, GCPNet outperformed its counterparts in terms of accuracy, precision, and F1 score (Table 2). According to the table, GCPNet achieved an accuracy of 97.2%, which is better

than other methods. In addition, the precision is about 5% higher than the second highest, ResNet50. The performance of the ordinary CNN is the worst because it does not use any special architectures or methods. Except for the recall, which is lower than ResNet50, all other indicators show that the performance of GCPNet is better than other methods, which also verifies that the model proposed in this experiment has excellent feasibility.

Table 2: Performance of models

Model	Accuracy	Precision	Recall	F1 score
CNN	0.927	0.824	0.929	0.873
VGG16	0.943	0.868	0.93	0.898
ResNet50	0.96	0.906	**0.952**	0.928
VGG16 + RF	0.937	0.846	0.935	0.888
ResNet50 + RF	0.928	0.829	0.923	0.874
GCPNet	**0.972**	**0.951**	0.946	**0.948**

Model performance was indicated by several metrics, and their definitions are provided as follows. Accuracy is indicated by the percentage of instances of true classification among all instances of classification, as written in (3).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

Precision is indicated by the percentage of instances of TPs among instances of positive classifications, as written in (4).

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

Recall is indicated by the percentage of positive cases correctly predicted by the model (as TPs), as written in (5). This is also called the sensitivity.

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

The F measure is a compound measure of precision and recall; it is defined as the harmonic average of both quantities, as written in (6). The F measure is dependent on a parameter β that ranges from 0 to positive infinity. The F measure is more reflective of precision and recall when β tends toward 0 and positive infinity, respectively.

$$F\ Measure = \left(1 + \beta^2\right) \times \frac{Precision \times Recall}{(\beta^2 * Precision) + Recall} \tag{6}$$

The F1 score is a special case of the F measure, in which $Beta = 1$, as written in (7).

$$F1\ score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{7}$$

4 Discussion

In this study, data was collected using X-ray machines commonly found in medical clinics. Chest X-ray images were obtained from patients at Guangzhou Medical Center, which were labeled by

medical personnel to ensure the accuracy of the prediction model and to prevent the machine learning model from learning incorrect information. The labeling personnel were required to have sufficient professional knowledge. In addition, to enable efficient training of the machine learning model, the number of pneumonia patients and healthy individuals in the data was balanced as much as possible. The training and testing data were collected and verified using the same process as described above. The feasibility of this method, which can effectively solve the problem of pneumonia diagnosis, was demonstrated by the research results. Once the machine learning model is trained and the chest X-ray is completed by the patient, the prediction model can provide recommendations to the physician. However, it is important to note that medical diagnostic support systems cannot fully represent the actual situation, and that a physician's expertise is necessary for deeper judgments in practice. By improving the efficiency and capacity of medical services, patients can be provided with more comprehensive medical care.

The dataset is originally segmented into ten parts in the process of cross-validation, and data augmentation is used for training data, rather than testing data, which prevents data leakage during the performance evaluation.

In order to maximize the probability of the center of an image belonging to the lungs, CNN strives to focus as much on it as possible. Therefore, if the CNN model's focus deviates from the center of a specific image, then we will retrain the list to ensure that the image is correctly identified if it strays from the center of the image.

As part of this research, a conventional X-ray machine is used to collect data, and TensorFlow is used to write the code as part of this project. For the purpose of measuring model efficiency, a cross-validation procedure was adopted to assess the accuracy of model identity in this paper. All datasets are gathered under the same conditions and processed in the same manner.

There are also many methods used in medical image recognition. These include applications that use chest X-rays to determine whether COVID-19 is diagnosed [31]. The reference also used models such as VGG19 for judgment, but this study also used data augmentation to improve the generalization of the model. In addition, there are also studies that analyze the degree of COVID-19 infection [32]. This reference also used transfer learning for judgment. In addition to using transfer learning, this study also used Grad-CAM to weight and improve the model for some feature maps for subsequent judgments, improving the accuracy of the judgment. The WE-layer ACP-based network (WACPN) [33] is also a neural network used for pneumonia diagnosis. The reference used the 2-dimensional wavelet entropy (2d-WE) layer and an adaptive chaotic particle swarm optimization (ACP) algorithm to train the neural network. The performance of this model was proven to be effective in the article. The use of optimization algorithms to train the neural network is different from this reference's traditional way of training neural networks and is also feasible.

This study found that the recognition performance of ResNet50 is superior. The reason is that the structure of this model can effectively overcome the problem of gradient disappearance. Through experiments, the conclusion is that compared with the model trained from scratch, transfer learning can achieve better results.

5 Conclusion

In this paper, A combination of a CNN model and four other pretrained models through transfer learning was proposed. In transfer learning, VGG16 and ResNet50 models act as feature extractors, and fully connected layers and RF act as classifiers. Data augmentation techniques are used in the

image preprocessing stage to prevent unbalance data. In evaluation tests, GCPNet with the ResNet50 extractor and fully connected layers performed the best. An integrated system that combines Grad-CAM and transfer learning was proposed to increase model performance. Grad-CAM is used to generate the heatmap, which shows the region that the model is focusing on. Images have an increased weight in the next iteration of training if the model did not focus on the right position. Future studies can implement different machine learning models, such as support vector machine and Adaboost. Additionally, different methods were designed to divide the heatmap of Grad-CAM to improve model accuracy.

Acknowledgement: We appreciate all the researchers for their contributions to this study.

Funding Statement: This work is supported by the National Science and Technology Council, Taiwan, under Grants NSTC 111-2218-E-194-007, NSTC 112-2218-E-194-006, MOST 111-2823-8-194-002, MOST 111-2221-E-194-052, MOST 109-2221-E-194-053-MY3, NSTC 112-2221-E-194-032. This work was financially partially supported by the Advanced Institute of Manufacturing with High-Tech Innovations (AIM-HI) from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan.

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Ping-Huan Kuo, Chia-Wei Jan; analysis and interpretation of results: Yu-Jhih Chiu, Kuan-Lin Chen; draft manuscript preparation: Chia-Wei Jan, Yu-Jhih Chiu, Kuan-Lin Chen, Ting-Chun Yao, Ping-Huan Kuo. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The used dataset can be downloaded from [29], as shown in the following website: https://data.mendeley.com/datasets/rscbjbr9sj/2.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] D. L. Duong, M. H. Kabir and R. F. Kuo, "Automated caries detection with smartphone color photography using machine learning," *Health Informatics Journal*, vol. 27, no. 2, pp. 146045822110075, 2021.

[2] Y. Omar, M. A. E. ElSheikh and R. Hodhod, "GLAUDIA: A predicative system for glaucoma diagnosis in mass scanning," *Health Informatics Journal*, vol. 27, no. 2, pp. 146045822110092, 2021.

[3] M. Intriago-Pazmino, J. Ibarra-Fiallo, J. Crespo and R. Alonso-Calvo, "Enhancing vessel visibility in fundus images to aid the diagnosis of retinopathy of prematurity," *Health Informatics Journal*, vol. 26, no. 4, pp. 2722–2736, 2020.

[4] I. Rudan, L. Tomaskovic, C. Boschi-Pinto and H. Campbell, "Global estimate of the incidence of clinical pneumonia among children under five years of age," *Bulletin World Health Organization*, vol. 82, pp. 895–903, 2004.

[5] E. Hosseini-Asl, J. M. Zurada, G. Gimel'farb and A. El-Baz, "3-D lung segmentation by incremental constrained nonnegative matrix factorization," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 5, pp. 952–963, 2016.

[6] S. Pereira, A. Pinto, V. Alves and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1240–1251, 2016.

[7] G. H. G. S. A. D. Dhanapala and S. Sotheeswaran, "Performance analysis for different optimizers on the CNN model for COVID-19 disease prediction based on chest X-ray images," in *2021 6th Int. Conf. on Information Technology Research (ICITR)*, Bangkok, Thailand, pp. 1–6, 2021.

[8] Y. Zhang, M. Attique Khan, Z. Zhu and S. Wang, "SNELM: SqueezeNet-guided ELM for COVID-19 recognition,"*Computer Systems Science and Engineering*, vol. 46, no. 1, pp. 13–26, 2023.

[9] J. Salamon and J. P. Bello, "Deep convolutional neural networks and data augmentation for environmental sound classification," *IEEE Signal Processing Letters*, vol. 24, no. 3, pp. 279–283, 2017.
[10] K. Kavin Kumar, P. M. Dinesh, P. Rayavel, L. Vijayaraja, R. Dhanasekar *et al.,* "Brain tumor identification using data augmentation and transfer learning approach," *Computer Systems Science and Engineering*, vol. 46, no. 2, pp. 1845–1861, 2023.
[11] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.
[12] L. Shao, F. Zhu and X. L. Li, "Transfer learning for visual categorization: A survey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 5, pp. 1019–1034, 2015.
[13] S. Niu, Y. Liu, J. Wang and H. Song, "A decade survey of transfer learning (2010–2020)," *IEEE Transactions on Artificial Intelligence*, vol. 1, no. 2, pp. 151–166, 2020.
[14] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu *et al.,* "A comprehensive survey on transfer learning," in *Proc. of IEEE*, vol. 109, no. 1, pp. 43–76, 2021.
[15] A. Shamsi, H. Asgharnezhad, S. S. Jokandan, A. Khosravi, P. M. Kebria *et al.,* "An uncertainty-aware transfer learning-based framework for COVID-19 diagnosis," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 4, pp. 1408–1417, 2021.
[16] S. Christodoulidis, M. Anthimopoulos, L. Ebner, A. Christe and S. Mougiakakou, "Multisource transfer learning with convolutional neural networks for lung pattern analysis," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 76–84, 2017.
[17] M. Lin, Q. Chen and S. Yan, "Network in network," arXiv1312.4400, 2013.
[18] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva and A. Torralba, "Learning deep features for discriminative localization," in *2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, Nevada, USA, pp. 2921–2929, 2016.
[19] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh *et al.,* "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *2017 IEEE Int. Conf. on Computer Vision (ICCV)*, Venice, Italy, pp. 618–626, 2017.
[20] G. Li, D. Xu, X. Cheng, L. Si and C. Zheng, "SimViT: Exploring a simple vision transformer with sliding windows," in *2022 IEEE Int. Conf. on Multimedia and Expo (ICME)*, Taipei, Taiwan, pp. 1–6, 2021.
[21] K. K. Singh and A. Singh, "Diagnosis of COVID-19 from chest X-ray images using wavelets-based depthwise convolution network," *Big Data Mining and Analytics*, vol. 4, no. 2, pp. 84–93, 2021.
[22] S. Basu, S. Mitra and N. Saha, "Deep learning for screening COVID-19 using chest X-ray images," in *2020 IEEE Symp. Series on Computational Intelligence (SSCI)*, Canberra, Australia, pp. 2521–2527, 2020.
[23] M. J. Hasan, M. S. Alom and M. S. Ali, "Deep learning based detection and segmentation of COVID-19 & pneumonia on chest X-ray image," in *2021 Int. Conf. on Information and Communication Technology for Sustainable Development (ICICT4SD)*, Dhaka, Bangladesh, pp. 210–214, 2021.
[24] D. Carmo, I. Campiotti, L. Rodrigues, I. Fantini, G. Pinheiro *et al.,* "Rapidly deploying a COVID-19 decision support system in one of the largest Brazilian hospitals," *Health Informatics Journal*, vol. 27, no. 3, pp. 146045822110330, 2021.
[25] S. Das, J. Nayak, S. Nayak and S. Dey, "Prediction of life insurance premium during pre- and post-COVID-19: A higher-order neural network approach," *Journal of the Institution of Engineers (India): Series B*, vol. 103, no. 5, pp. 1747–1773, 2022.
[26] A. S. Shaik, R. K. Karsh, M. Islam and S. P. Singh, "A secure and robust autoencoder-based perceptual image hashing for image authentication," *Wireless Communications and Mobile Computing*, vol. 2022, pp. 1–17, 2022.
[27] A. S. Shaik, R. K. Karsh, M. Suresh and V. K. Gunjan, "LWT-DCT based image hashing for tampering localization via blind geometric correction," in *ICDSMLA 2020*, Pune, India, pp. 1651–1663, 2022.
[28] A. S. Shaik, R. K. Karsh, M. Islam and R. H. Laskar, "A review of hashing based image authentication techniques," *Multimedia Tools and Applications*, vol. 81, no. 2, pp. 2489–2516, 2022.
[29] D. Kermany, K. Zhang and M. Goldbaum, "Labeled optical coherence tomography (OCT) and chest X-ray images for classification," *Mendeley Data*, vol. 2, no. 2, 2018.

[30] D. S. Kermany, M. Goldbaum, W. Cai, C. C. S. Valentim, H. Liang *et al.,* "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131.e9, 2018.
[31] V. Narasimha and D. M. Dhanalakshmi, "Detection and severity identification of COVID-19 in chest X-ray images using deep learning," *International Journal of Electrical and Electronics Research*, vol. 10, no. 2, pp. 364–369, 2022.
[32] S. S. Skandha, L. Saba, S. K. Gupta, V. K. Kumar, A. M. Johri *et al.,* "Characterization of COVID-19 severity in infected lungs via artificial intelligence transfer learning," in *Multimodality Imaging*, vol. 1. Amsterdam, Netherlands: IOP Publishing, pp. 12–1–12–25, 2022.
[33] S. H. Wang, M. Attique Khan, Z. Zhu and Y. D. Zhang, "WACPN: A neural network for pneumonia diagnosis," *Computer Systems Science and Engineering*, vol. 45, no. 1, pp. 21–34, 2023.

Appendix A

A.1 Convolution Neural Network

A CNN is used to create the model in this paper. A CNN is a neural network model that can extract the features of images and speech. The foundations of a CNN are a convolution layer, a pooling layer, and a fully connected layer.

The convolution layer uses feature detectors (filters) to perform matrix multiplication with the input image, generating feature maps. The pooling layer primarily uses max pooling. Max pooling can be used to select the maximum value in the matrix by the pool size. Finally, a fully connected layer flattens the previous result and connects it to a basic neural network.

Fig. 19 shows the structure of the CNN designed in this paper. Two consecutive convolution layers and one max pooling layer are used for each block. The fourth block is connected to dense layers. However, the input of the dense layer is one dimensional, and the output of convolution layers is not in the same dimension as the dense layer. Therefore, the flatten layer must be used to transform the two layers into one dimension. Finally, to prevent overfitting, dropout layers were added to remove some neurons.

This model uses a Rectified Linear Unit (ReLU) function as an activation function. Furthermore, the last layer is composed of one neuron unit configured with a binary classifier sigmoid function.

A.2 Random Forest

Random forest (RF) is an ensemble learning model using bagging as an ensemble method and a decision tree as the base model. The advantage of the bagging algorithm lies in its ability to be used to process the noise data (bad data) present in the original training samples. Using bagging screening can potentially prevent noise data from being chosen and reduce the instability of the model. The structure of RF is constructed using several decision trees and the output of each tree determines the final prediction. Finally, the majority voting method is used to obtain the final prediction of the model. Fig. 20 shows the RF.

After the model outputs a prediction after training, the model is evaluated by comparing the prediction against the ground truth. For the classification, a confusion matrix was used to evaluate the result.

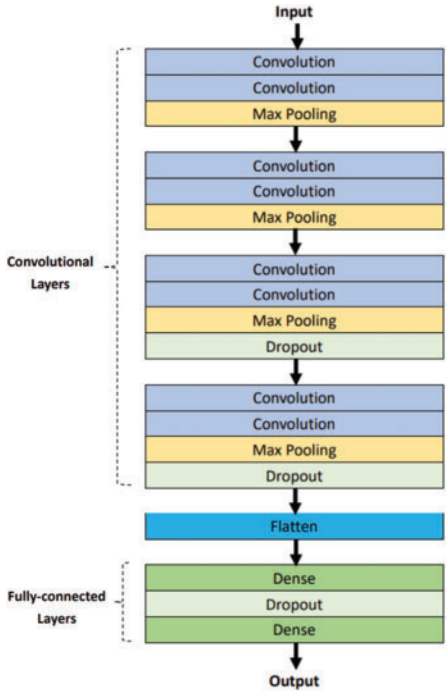


Figure 19: Structure of conventional CNN model

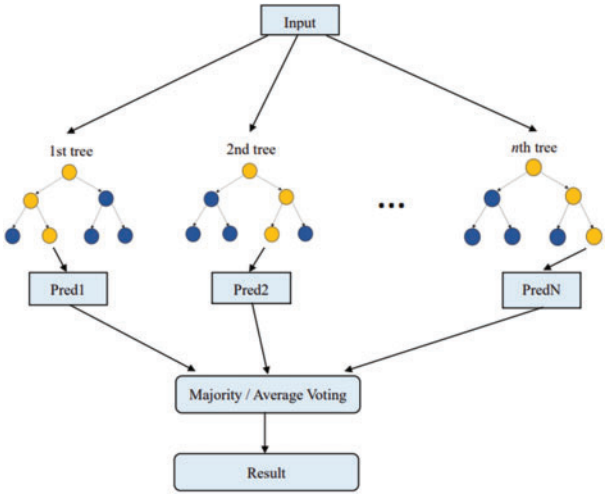


Figure 20: Random forest

A.3 Confusion Matrix

A confusion matrix is a data visualization tool. For n categories, the confusion matrix has the form of an $n \times n$ table showing n^2 parameters, including the number of True Positives (TPs), False Positives (FPs), False Negatives (FNs), and True Negatives (TNs). Positive and negative cases were defined to be normal and pneumonia cases, respectively. The definition of confusion matrix is shown in Fig. 21. The confusion matrixes results of the models are shown in Fig. 22.

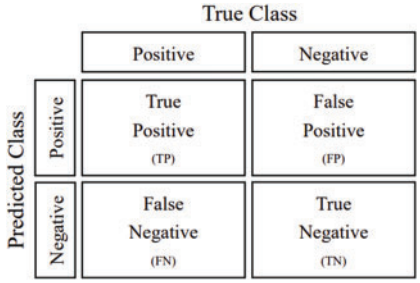


Figure 21: Confusion matrix

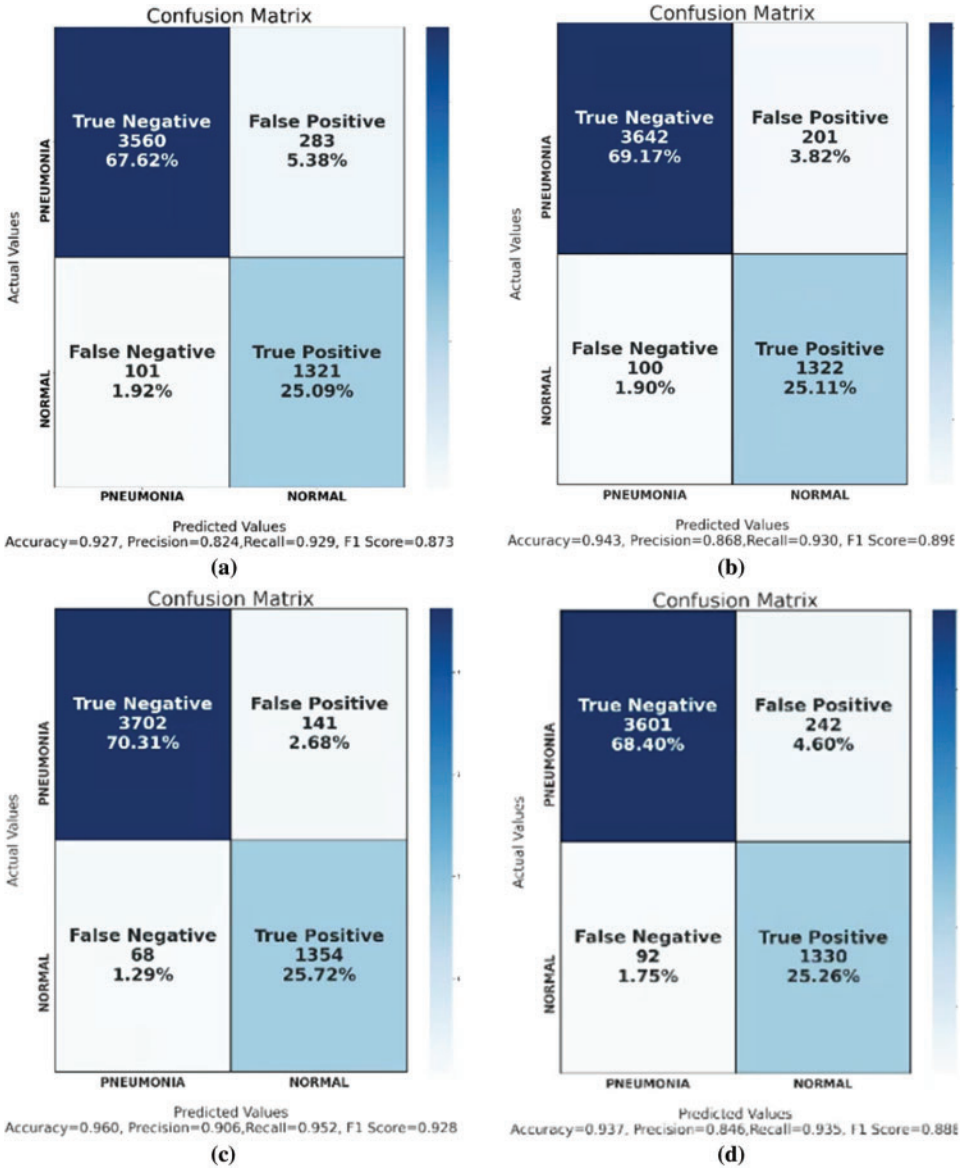


Figure 22: (Continued)

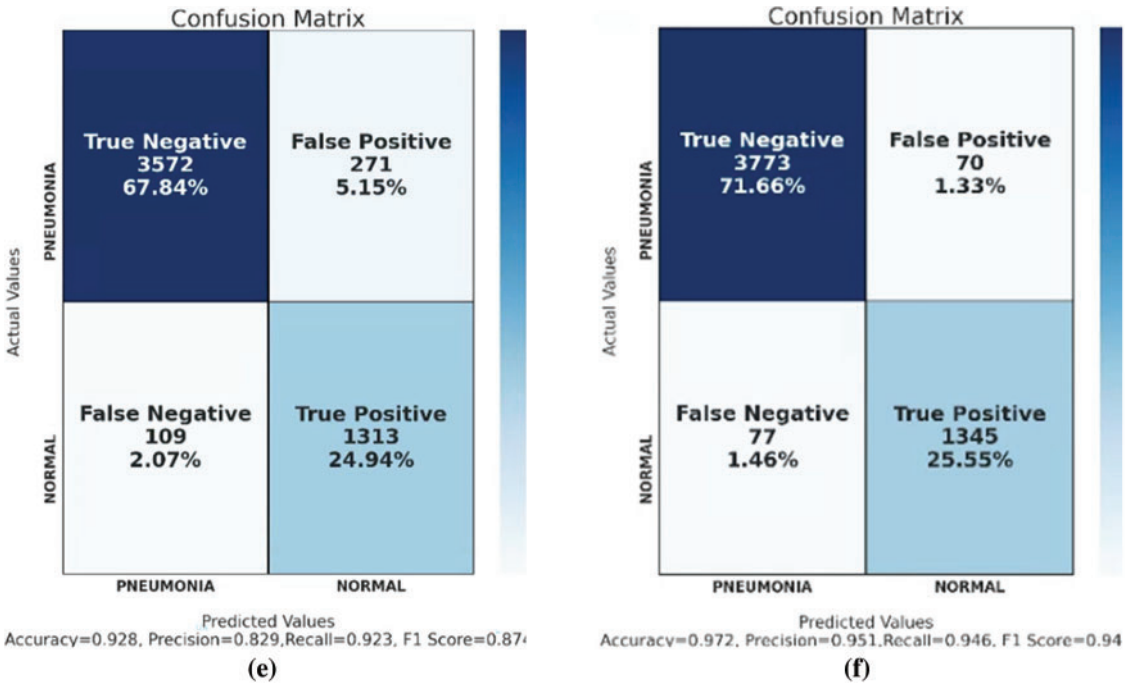


Figure 22: Confusion matrix of (a) CNN, (b) VGG16, (c) ResNet50, (d) VGG16 + RF, (e) ResNet50 + RF, and (f) GCPNet