AN INVESTIGATION INTO DETECTING PNEUMONIA THROUGH IMAGE PROCESSING AND OBJECT DETECTION

Prithvi Sairaj Krishnan

Department of Computer Science, Westwood High School, Austin, United States of America

ABSTRACT

One of the most common respiratory infections that causes substantial morbidity and mortality worldwideis pneumonia, particularly in poorer countries with poor medical infrastructure. Chest X-ray imaging is essential for early diagnosis, although it can be difficult. In order to identify pneumonia from chest X-rays, this study created an automated deep learning computer-aided diagnosis method. Three pre-trained convolutional neural network models (ResNet-18, DenseNet-121), together with a newly developed weighted average ensemble approach based on evaluation metric scores, were used in the ensemble. Tested using five-fold cross-validation on two public X-ray datasets for pneumonia, the methodoutperformed state-of-the-art techniques with high accuracy (98.2%, 86.7%) and sensitivity (98.19%, 86.62%). Over 2.5 million fatalities globally are attributed to pneumonia each year. This precise automated model can help radiologists diagnose patients in a timely manner, particularly in situations with limited resources. How it is included into clinical decision assistance systems has the potential to improve pneumonia management and outcomes significantly.

KEYWORDS

Convolutional Neural Networks, Pneumonia, Infection, X-Rays, Model, Machine Learning

1. INTRODUCTION

Pneumonia is a serious lung illness caused by bacteria, viruses, or fungus. Pneumonia can lead to pleural effusion, a disorder characterized by fluid collection and inflammation of the air sacs in the lungs. Pneumonia is a major cause of mortality for children under five, particularly in developing and growing countries where there is a high pollution rate, overcrowding, poor hygiene, and limited access to healthcare. For pneumonia to be treated effectively and prevent death, it must be detected early. Radiological tests like computed tomography (CT), magnetic resonance imaging (MRI), or X-rays are commonly used to detect pneumonia. X-ray imaging isa fairly cost, non-invasive method of assessing the lungs. Infiltrates are white areas that are shown by red arrows in the sample image, they distinguish a pneumonic condition from a healthy lung. However, chest X-ray examinations for pneumonia detection are subject tosubjective variability. Therefore, an automated system for pneumonia detection is necessary. In this study, the researcher developed a computer-aided diagnosis (CAD) system that utilises an ensemble of deep transfer learning models for the accurate classification of chest X-ray images to detect pneumonia.

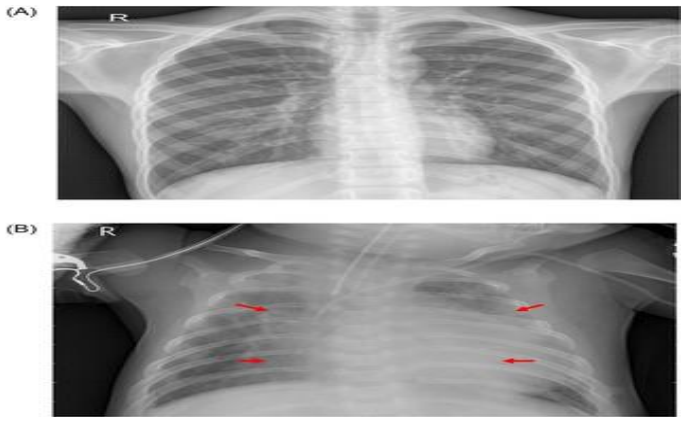

Fig 1. Examples of two X-ray plates that display (a) a healthy lung and (b) a pneumonic lung.

The red arrows in (b) indicate white infiltrates, a distinguishing feature of pneumonia. The images were taken from the Kermany dataset [2].

Convolutional neural networks (CNNs), in particular, are potent artificial intelligence tools frequently employed in deep learning to solve challenging computer vision problems. But for these models to function at their best, a lot of data is needed, and this can be difficult to come by for biomedical image classification tasks because each image must be classified by a team of highly qualified clinicians, which is costly and time-consuming. One method to overcome this challenge is transfer learning, which involves taking a model that was trained on a massive dataset—like ImageNet, which has over 14 million images—and using the learned network weights to solve a problem and make accurate predictions a final prediction for a test sample by combining the decisions of numerous classifiers is a popular approach known as ensemble learning. It seeks to extract the discriminative information from every base classifier to produce more accurate predictions. Average probability, weighted average probability, and majority voting are examples of common ensemble approaches. Although the average probability-based ensemble gives every basic learner equal weight, it is a better idea to give the base classifiers weights because some may be better at capturing information than others for a given task. Nonetheless, it guarantees improved performance, it is essential to ascertain the ideal weight values for every classifier.

For this work, I devised a unique weight allocation technique based on four assessment metrics: accuracy, recall, f1-score, and area under the receiver operating characteristic (ROC) curve (AUC). The ideal weights were assigned to three basic CNN models: GoogLeNet, ResNet-18, and DenseNet-121. Previous studies have largely concentrated on classification accuracy when figuring the base learner weights, which might not be enough, particularly when working with datasets that aren't dispersed uniformly throughout the class. Furthermore, other criteria might provide more relevant information for prioritizing the basic learners. The full process of the recommended ensemble structure is shown in Figure 2.

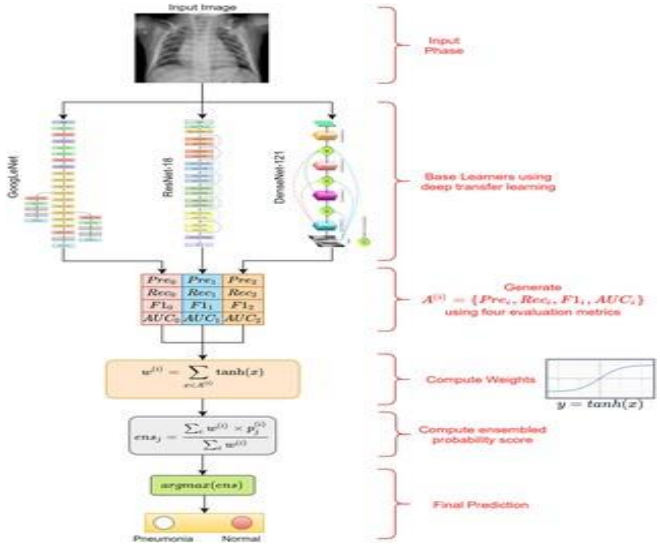

Fig 2. Representation of the proposed pneumonia detection framework.

Pre = Precision score, Rec = Recall score, F1 = F1-score, AUC = AUC score, and A(i) = {Prei, Reci, F1i, AUCi}; w(i) is the weight generated for the i^{th} base learner to compute the ensemble, is the probability score for the j^{th} sample by the i^{th} classifier, and enjoy is the fused probability score for the j^{th} sample; and the arg max function returns the position having the highest value in a 1D array, i.eIn this case, it generates the predicted class of the sample.

2. RELATED WORK

Table 1. Existing methods for pneumonia detection

Method	Approach	Merits	Demerits
Alhakbi et al. [9]	• Transfer Learning using InceptionResNet-V2	Reuse of models pretrained on a large dataset	Overamplified for a complex pattern recognition task; Performance obtained is poor and not fit for practical use
Rahman et al. [10]	• Transfer Learning using DenseNet-201		
Liang et al. [11]	• Transfer learning using ResNet-50 pre-trained on ChestX-ray14 dataset		
Ibrahim et al. [12]	• Transfer learning using AlexNet		
Zulbeir et al. [13]	• Transfer learning using VGG-16		
Rajpurkar et al. [14]	• Transfer learning using DenseNet-121		
Alhakbi et al. [15]	• Used generative adversarial networks to generate synthetic data. • Classification using ResNet-152	Generation of synthetic data to balance the classes of the data because medical data are scarce	Classification results (81% accuracy rate) are not fit for practical use
Chandra et al. [16]	• Segmentation of lung X-rays using image processing • Extraction and classification of eight statistical features	Segmentation of lungs before classification allows localization of the disease	The use of handcrafted features limits its ability to perform in complex pattern recognition tasks; Evaluation on a small dataset (412 images) cannot be generalized
Kuo et al. [17]	• Used 13 features from patient data to fit traditional classifiers	Use of 10-fold cross validation with 3 repeats avoids overfitting	Patient data are often private and not publicly available to fit to classification models
Yue et al. [18]	• Segmented lung lobes using U-Net • Extracted and classified radiomic features from CT scan images	Segmentation before classification helps extract important features for radiologists and allows localization of the disease	Method evaluated on a small dataset (72 lesion segments); and thus difficult to generalize
Sharma et al. [19]	• Devised a CNN model for classification of X-ray images	Automatic feature learning for complex tasks	Simple linearly progressing CNN model increases computation cost without providing strong boost to performance
Stephen et al. [20]	• Developed a simple seven-layer CNN model for classification of X-ray images.		
Ianizek et al. [21]	• Developed a deep learning framework based on adversarial optimization	Adversarial optimization removed dependency on the source of the dataset and view of the X-rays for classification	Results (AUC 74.5%) are not fit for deployment in the field
Zhang et al. [22]	• Developed a confidence-aware module for anomaly detection in lung X-ray images	Posing the detection task as a one-class problem helped improve the model performance	The sensitivity obtained on the dataset was too low (73.50%) for practical use
Tsiknakis et al. [23]	• Applied fuzzy rose transformation to X-ray images • Extracted local features for classification using an ensemble of traditional classifiers	Generation of three different feature images improves the model performance	Handcrafted feature extraction limits performance in complex pattern recognition tasks; Evaluation on a small dataset cannot be generalized
Jaiswal et al. [24]	• Developed a mask region-based CNN for segmentation • Used an ensemble model for image thresholding	Use of threshold value in background boosts the performance	An irregular trend was observed, where results of the training set were lower than those of the testing set
Gabruseva et al. [25]	• Localized pulmonary opacity based on a single shot detector • Used a snapshot ensemble model for segmentation	One shot detector alleviates the problem of scarcity of data.	Irregular trend of validation loss over epochs during model training
Pan et al. [26]	• Used an ensemble of Inception-ResNet v2, XceptionNet, and DenseNet-169 for bounding box prediction	Ensemble learning allows the fusion of salient properties of all its base learners	Pan et al. [26] suspect that their model evaluated on only one dataset may not generalize over data acquired from a different source

3. MOTIVATIONS AND CONTRIBUTIONS

Many people, particularly children, suffer greatly from pneumonia. This condition is most common in developing and impoverished nations when risk factors such as overcrowding, poor hygiene, hunger, and a lack of proper medical facilities are present. It takes an early diagnosis to fully recover from pneumonia. The most popular diagnostic technique is X-ray examination, however, it depends on the radiology's interpretive skills, which frequently causes radiologists to disagree. For an accurate diagnosis, then, a generalisation-capable automated computer-aided diagnosis (CAD) system is required. The majority of earlier research ignored the possible advantages of ensemble learning in favor of creating a single convolutional neural network (CNN) model for the categorization of pneumonia. Better predictions are made possible by ensemble learning, which combines discriminative data from several base learners. Ensemble learning was used in this study to address a lack of medical data by using transfer learning models as base learners and ensembling their decision scores.

By using a weighted average ensemble approach, an ensemble framework was created to improve the performance of basic CNN learners in the categorization of pneumonia cases. Rather than relying just on classifier performance or experimental results, the weights assigned to the classifiers were determined by integrating four assessment metrics: precision, recall, f1-score, and area under the curve (AUC), using a hyperbolic tangent function. The RSNA Pneumonia Detection Challenge dataset and the Kermany dataset, two publically available chest X-ray datasets, were used to evaluate the proposed model using five-fold cross-validation. The results outperformed the state-of-the-art methods, suggesting that the method may be applied in practical settings.

4. PROPOSED METHOD

In this study, I designed an ensemble framework consisting of three classifiers: GoogLeNet, ResNet-18, and DenseNet-121, using a weighted average ensemble scheme. The weights allocated to the classifiers were generated using a novel scheme, as explained in detail below.

4.1. Googlenet

The GoogLeNet design, which was proposed by Szegedy et al., is a 22-layer deep network that employs "inception modules" as opposed to layers that are uniformly progressive. An inception block may accommodate many units at each level by supporting parallel convolution and pooling layers. Nevertheless, the extra parameters lead to an unmanageable computing complexity. The GoogLeNet model employs inception blocks with dimension reduction, as shown in Fig. 3(b), to control the computational complexity as opposed to the naïve inception block (Fig. 3(a)) used in previous work. GoogLeNet's performance, which introduced the inception block, shows that an optimal sparse architecture made from easily obtainable dense building blocks improves the performance of artificial neural networks for computer vision applications. Design of the GoogLeNet model shown in Fig. 4.

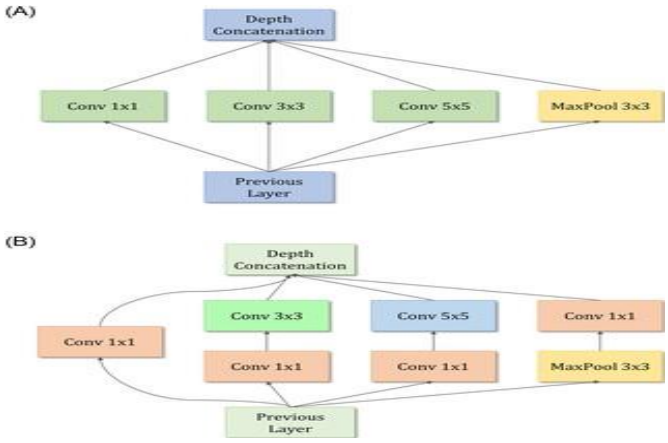

Fig 3. Inception modules in the GoogLeNet architecture.

(a) The naive inception block is replaced by (b) the dimension reduction inception block in the GoogLeNet architecture to improve computational efficiency.

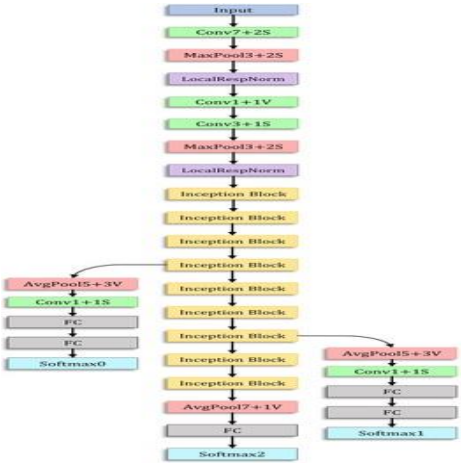

Fig 4. The architecture of the GoogLeNet model was used in the study

The inception block is shown in Fig 3(b).

4.2. Resnet-18

Deep network training is made more successful by Huang et al.'s ResNet-18 model, which is based on a residual learning methodology. The residual blocks of ResNet models aid in network optimization, improving model accuracy overall. This is distinct from the initial unreferenced mapping present in inversely continuing convolutions. These residuals, or links, provide identity mapping without adding parameters or increasing computing complexity. The architecture of the ResNet-18 model is shown in Figure 5.

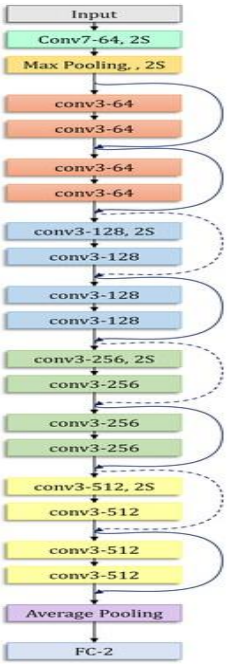

Fig 5. The architecture of the ResNet-18 model used in this study

4.3. Densenet-121

According to Huang et al., DenseNet topologies provide a rich feature representation and are computationally efficient. The primary rationale is because each layer of the DenseNet model's feature maps are concatenated with feature maps from all preceding levels, as seen in Fig. 6. Because the convolutional layers can accommodate fewer channels, the model becomes computationally efficient when the number of trainable parameters decreases. Concatenating the feature maps from previous layers with the current layer further enhances the feature representation capacity.

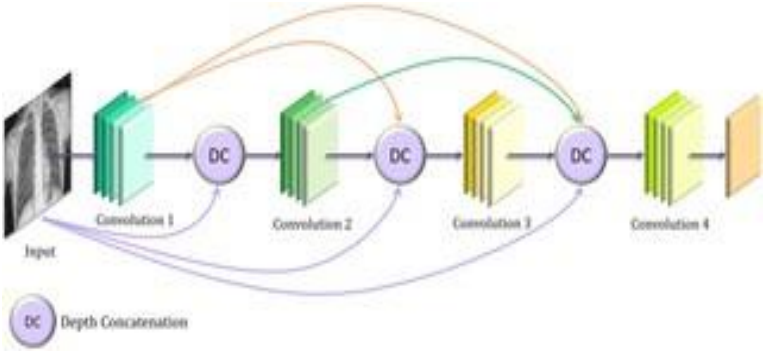

Fig 6. The basic architecture of the DenseNet convolutional neural network model.

The values of the hyperparameters used for training the learning algorithms (base learners) were set empirically and are shown in Table 2.

Table 2. Hyperparameters are used for training the convolutional neural network base learners.

Hyperparameter	Value
Optimizer	Adam
Loss Function	Cross Entropy
Initial Learning Rate	0.0001
Learning Rate Scheduler	ReduceLROnPlateau
No. of Epochs	30

5. PROPOSED ENSEMBLE SCHEME

Better predictions than any of its base learners are produced by the ensemble learning model, which assists in incorporating the discriminative information from all of its constituent models. The weighted average ensemble is an effective method for classifier fusion. But one of the most important factors in guaranteeing the ensemble's success is the selection of weights given to the corresponding base learners. The majority of methods in the literature either experiment or just consider the accuracy of the classifier when determining the weights. If there is a class imbalance in the dataset, this method might not be appropriate. Other assessment metrics, such f1-score, area under the curve (AUC), recall (sensitivity), and accuracy, may offer more reliable data for establishing the base learners' priority. This study came up with a unique plan to achieve this goal for weight allocation, which is explained below.

First, the probability scores obtained during the training phase by the base learners are utilised to calculate the weights assigned to each base learner using the proposed strategy. These generated weights are used in the formation of an ensemble trained on the test set. This strategy is implemented to ensure that the test set remains independent for predictions. The predictions of the i^{th} model are generated and compared with the true labels (y) to generate the corresponding precision score (pre^i), recall score (rec^i), f1-score ($f1^i$), and AUC score (AUC^i). Assume that this forms an array $A^i = \{pre^i, rec^i, f1^i, AUC^i\}$. The weight ($w^i$) assigned to each classifier is then computed using the hyperbolic tangent function, as shown in Eq 1. The range of the hyperbolic tangent function is [0, 0.762] because x represents an evaluation metric, the value of which is in the range [0, 1]. It monotonically increases in this range; thus, if the value of a metric x is high, the tanh function rewards it by assigning it a high priority; otherwise, the function penalises it.

These weights (w(i)) computed by Eq 1 are multiplied by the decision scores of the corresponding base learners to compute the weighted average probability ensemble, as shown in Eq 2, where the probability array (for a binary class dataset) of the j^{th} test sample by the i^{th} base classifier is, where $a \leq 1$ and the ensemble probability for the sample is $ensemble\_prob_j = \{b, 1 - b\}$. Finally, the class predicted by the ensemble is computed by Eq 3, where $prediction_j$ denotes the predicted class of the sample.

6. RESULTS AND DISCUSSION

This section displays the evaluation results of the recommended approach. I utilized two freely available datasets of chest X-rays for pneumonia. 5856 chest X-ray pictures that are unequally distributed between the "Normal" and "Pneumonia" classifications make up the first dataset, known as the Kermany dataset. The images feature a diverse range of people and kids. The second dataset was released as a Kaggle challenge for pneumonia detection and made available

by the RSNA. Table 3 displays the distribution of images between the two datasets as well as the picture descriptions for the training and testing sets for each fold of the five-fold cross-validation approach employed in this investigation. Additionally, the implications of the obtained results are discussed. A comparative examination was done to show how much better the proposed method is over other models and frequently used ensemble techniques published in the literature.

Table 3. Description of images in the training and testing sets in each fold of five-fold cross-validation in the two datasets used in this study.

Dataset	Division	Class	No. of Images	Size of Images (Range)	Size of Resized Images
Kermany	Train	Normal	1267	[127×384×3]–[2713×2517×3]	[224×224×3]
		Pneumonia	3419		
	Test	Normal	316	[189×498×3]–[2458×2720×3]	[224×224×3]
		Pneumonia	854		
	Total Images			5856	
RSNA	Train	Lung Opacity	16480	[1024×1024×3]	[224×224×3]
		No Lung Opacity	8851		
	Test	Lung Opacity	6111	[1024×1024×3]	[224×224×3]
		No Lung Opacity	1361		
	Total Images			36681	

7. EVALUATION METRICS

Four common evaluation measures were applied to the two pneumonia datasets to assess the suggested ensemble method: f1-score (F1), accuracy (Acc), precision (Pre), and recall (Rec). First, I define the phrases "True Positive," "False Positive," "True Negative," and "False Negative" to define these evaluation measures. Now, the four evaluation metrics can be defined as:

$$Acc = \frac{TP + TN}{TP + FP + TN + FN}$$

$$Pre = \frac{TP}{TP + FP} \quad Rec\ (or\ Sensitivity) = \frac{TP}{TP + FN}$$

$$F1 = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}}$$

Fig 7. The different evaluation metrics of the pneumonia detection ensemble model using components of the confusion matrix make up the evaluation metrics.

The accuracy rate provides a broad idea of the proportion of the model's predictions that were realized. A model's high accuracy rate does not, however, imply that it can differentiate between several classes equally if the dataset is imbalanced. More specifically, medical image categorization requires a globally applicable model. In these cases, looking at the "precision" and "recall" variables will help you understand how well the model performs. The accuracy of the positive label prediction made by the model is displayed. This is the ratio between all of the model's predictions and the accurate forecasts. Conversely, "recall" measures how much of the positive ground truth data the model correctly predicted. FN and FP quantities can be reduced by the model based on these two evaluation criteria. The accuracy rate provides a broad idea of the proportion of the model's predictions that were realized. A model's high accuracy rate does not, however, imply that it can differentiate between several classes equally if the dataset is imbalanced. More specifically, medical image categorization requires a globally applicable

model. In these cases, looking at the "precision" and "recall" variables will help you understand how well the model performs. The accuracy of the positive label prediction made by the model is displayed. This is the ratio between all of the model's predictions and the accurate forecasts. Conversely, "recall" measures how much of the positive ground truth data the model correctly predicted. FN and FP quantities can be reduced by the model based on these two evaluation criteria.

8. IMPLEMENTATION

This work employed a 5-fold cross-validation technique to evaluate the performance of the proposed ensemble model in detail. Tables 4 and 5 display the findings for the RSNA challenge dataset and the Kermany dataset, respectively, along with the average and standard deviation values for all five folds. The outstanding accuracy and sensitivity (recall) ratings show the reliability of the recommended strategy. Additionally, Figures 7 and 8 display the confusion matrices on the RSNA and Kermany datasets, and Figure 8 displays the ROC curves generated by the recommended approach for each of the two datasets' five cross-validation folds.

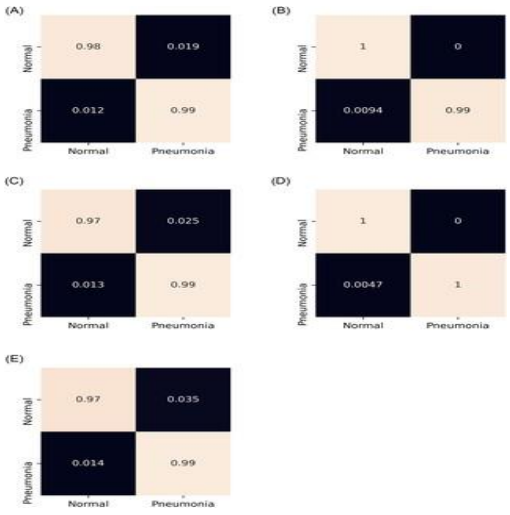

Fig 8. Confusion matrices were obtained on the Kermany pneumonia chest X-ray dataset by the proposed method by 5-fold cross-validation. a) Fold-1. (b) Fold-2. (c) Fold-3. (d) Fold-4. (e) Fold-5.

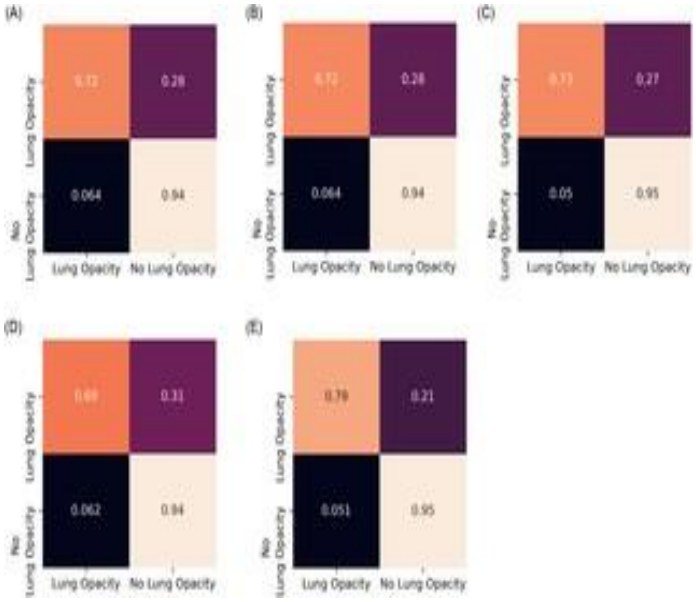

Fig 9. Confusion matrices obtained on the Radiological Society of North America pneumoniachallenge chest X-ray dataset by the proposed method by five-fold cross-validation. a) Fold-1. (b) Fold-2. (c) Fold-3. (d) Fold-4. (e) Fold-5.

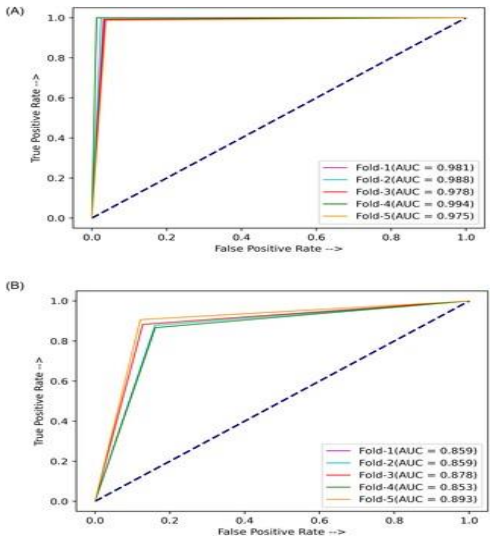

Fig 10. Receiver operating characteristic curves obtained by the proposed ensemble method on the two pneumonia chest X-ray datasets used in this research:(a) Kermany dataset [2]. (b) RSNA challenge dataset [16].

Table 4. Results of five-fold cross-validation of the proposed ensemble method on the pneumonia Kermany dataset [2].

Fold	Acc(%)	Pre(%)	Rec(%)	F1(%)	AUC(%)
1	98.63	98.64	98.63	98.63	98.12
2	99.31	99.33	99.32	99.32	98.82
3	98.38	98.46	98.38	98.29	97.86
4	99.68	99.66	99.66	99.66	99.43
5	98.03	98.03	98.03	98.03	97.54
Avg±Std. Dev.	**98.81±0.61**	**98.82±0.59**	**98.80±0.60**	**98.79±0.61**	**98.35±0.68**

Avg: average *Std.Dev*: Standard Deviation.

Table 5. Results of five-fold cross-validation of the proposed ensemble method on the pneumonia Radiological Society of North America challenge dataset.

Fold	Acc(%)	Pre(%)	Rec(%)	F1(%)	AUC (%)
1	86.63	86.78	86.63	86.70	86.63
2	86.78	87.05	87.05	87.05	86.78
3	87.97	88.00	87.80	87.90	87.97
4	85.98	86.00	86.63	86.31	85.98
5	86.89	86.63	86.98	86.80	86.89
Avg±Std. Dev.	**86.85±0.72**	**86.89±0.73**	**87.02±0.48**	**86.95±0.59**	**86.85±0.72**

Avg: average *Std.Dev*: Standard Deviation.

Figure 10 showcases the accuracy rates achieved by the base learners in transfer learning using different optimizers on the Kermany dataset. The Adam optimizer yielded the best results for all three base learners and was consequently chosen as the optimizer for training the base learners in the ensemble framework.

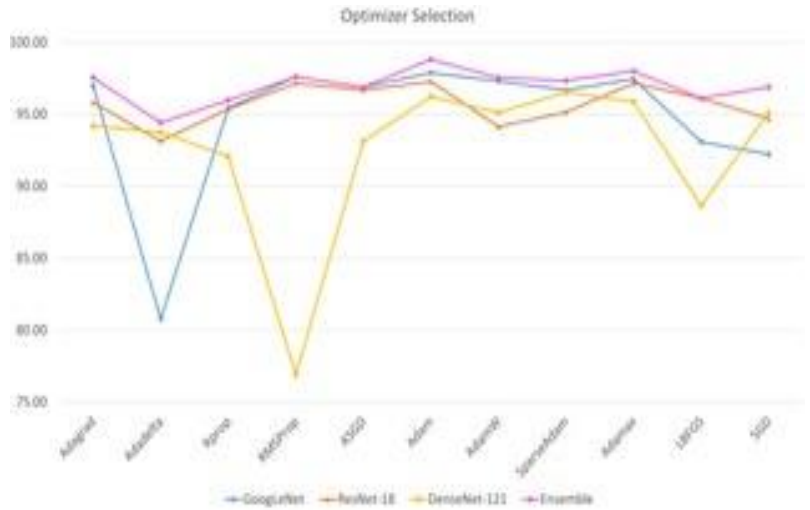

Fig 11. Variation of accuracy rates on the Kermany dataset [2]) was achieved by the three base learners, GoogLeNet, ResNet-18, and DenseNet-121, and their ensemble, according to the optimizers chosen for fine-tuning.

The results of many ensembles using three different base learners are displayed in Table 6, which also includes newly proposed architectures on the Kermany dataset: GoogLeNet, ResNetvariants, DenseNet variations, MobileNet v2, and NASMobileNet. The results validate the choice of base learner combinations used in this study, namely GoogLeNet, ResNet-18, and DenseNet-121. This ensemble combination has a 98.2% accuracy rating. The ensemble consisting of GoogLeNet, ResNet-18, and MobileNet v2 yielded the second-best result, with anaccuracy rate of 98.54%. In addition, the models were trained to find the optimal configuration once a few layers were fixed for the chosen group of base learners. The findings, which are shown in Figure 12, demonstrate that the ensemble worked best on both datasets when every layer was trainable (0 layers frozen). Because of this, the precise setting was chosen for the ensemble framework.

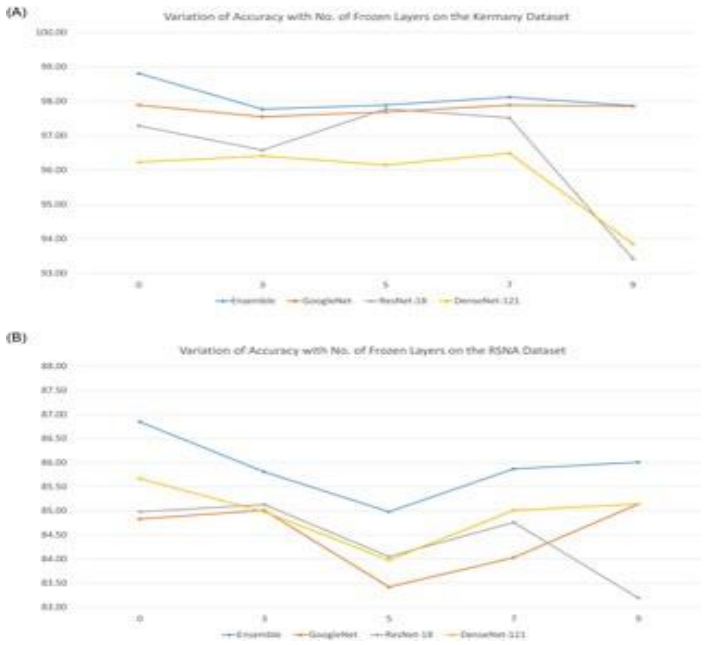

Fig 12. Variation in performance (accuracy rates) of the ensemble concerning the number of fixed non-trainable layers in the base learners on the two datasets used in this study:(a) Kermany dataset [2]. (b) RSNA challenge dataset [16].

Table 6. Results of extensive experiments performed to determine the base learners for forming the ensemble in this study.

Model-1	Model-2	Model-3	Acc(%)	Pre(%)	Rec(%)	F1(%)
NasNetMobile	MobileNet v2	ResNet-152	96.67	96.70	96.67	96.68
NasNetMobile	MobileNet v2	ResNet-50	97.00	97.02	97.01	97.01
NasNetMobile	MobileNet v2	DenseNet-169	96.41	96.40	96.41	96.41
NasNetMobile	MobileNet v2	DenseNet-201	96.06	96.21	96.07	96.11
MobileNet v2	ResNet-152	DenseNet-169	96.92	97.03	96.92	96.95
MobileNet v2	ResNet-50	DenseNet-169	97.77	97.82	97.78	97.79
MobileNet v2	ResNet-50	DenseNet-201	95.98	96.40	95.98	96.06
MobileNet v2	ResNet-152	DenseNet-201	94.87	95.57	94.87	94.99
NasNetMobile	ResNet-152	DenseNet-169	95.21	95.71	95.21	95.31
NasNetMobile	ResNet-152	DenseNet-201	92.56	94.06	92.56	92.81
NasNetMobile	ResNet-50	DenseNet-169	96.41	96.66	96.41	96.46
NasNetMobile	ResNet-50	DenseNet-201	92.99	94.28	92.99	93.20
GoogLeNet	ResNet-152	DenseNet-121	97.17	97.37	97.18	97.21
GoogLeNet	ResNet-152	DenseNet-201	93.04	95.65	95.04	95.15
GoogLeNet	ResNet-18	DenseNet-201	98.20	98.23	98.21	98.21
GoogLeNet	MobileNet v2	DenseNet-121	98.29	98.29	98.29	98.29
GoogLeNet	ResNet-18	MobileNet v2	98.54	98.54	98.55	98.54
GoogLeNet	MobileNet v2	NasNetMobile	98.12	98.13	98.12	98.11
GoogLeNet	**ResNet-18**	**DenseNet-121**	**98.81**	**98.82**	**98.80**	**98.35**

8.1. Comparison with State-of-the-Art Methods

On the Kermany pneumonia dataset, Table 7 presents a performance comparison between the suggested ensemble framework and the approaches that are already available in the literature. It should be mentioned that the suggested approach performed better than any other approach.
Moreover, it is noteworthy that the proposed ensemble framework outperformed all of the previous methods (Mahmud et al. [18], Zubair et al. [8], Stephen et al. [15], Sharma et al. [14], Liang et al. [6]) that relied on using a single CNN model for the classification of pneumonic lung X-ray images. This suggests that the ensemble technique developed in this study is a dependable method for the image classification task at hand. As far as we are aware, no research has been done on the categorization of pictures in the RSNA pneumonia dataset exist. Hence, for this dataset, I compared the performance of the proposed model to that of several baseline CNN models.

Table 7. Comparison of the proposed method with other methods in the literature on the Kermany pneumonia dataset [2] and the Radiological Society of North America challenge dataset [16].

Dataset	Method	Acc(%)	Pre(%)	Rec(%)	F1(%)	AUC(%)
Kermany	Mahmud et al. [39]	98.10	98.00	98.50	98.80	-
	Zubair et al. [13]	96.60	97.20	98.10	97.65	-
	Stephen et al. [20]	93.73	-	-	-	-
	Sharma et al. [19]	90.68	-	-	-	-
	Liang et al. [11]	90.30	89.10	96.70	92.70	-
	Proposed Method	**98.81**	**98.82**	**98.80**	**98.79**	**98.35**
RSNA	Antin et al. [40]	-	-	-	-	61.00
	Zhou et al. [41]	79.70	-	-	80.60	-
	Yao et al. [12]	-	-	-	-	71.30
	Rajpukar et al. [14]					76.80
	Proposed Method	**86.85**	**86.89**	**87.02**	**86.95**	**86.85**

Table 8 compares the assessment results of the proposed technique with those of the basic CNN models used to construct the ensemble and various other conventional CNN transfer learning models on both datasets utilized in this work. On both datasets, it is evident that the suggested ensemble approach did rather well compared to alternative transfer learning models and the base learners. In addition, Table 9 compiles the findings to demonstrate the superiority of the suggested ensemble scheme over conventional popular ensemble strategies. For both the Kermany and RSNA challenge datasets, the average results across the five folds of cross-validation are displayed. The ensembles employed the same three basic CNN learners, GoogLeNet, ResNet-18, and DenseNet-121. Popular ensemble techniques were outperformed by the suggested ensemble approach. It is evident from both datasets that the weighted average ensemble that uses the accuracy metric as the sole weighting factor produced the best results, nearly matching the suggested ensemble approach. The class that received the most votes from the base learners is expected to be the sample class in the majority voting-based ensemble. In the maximum probability ensemble, all base learners' probability scores are added up, and the class with the highest probability is designated as the sample's predicted class. In the average probability ensemble, on the other hand, each contributing classifier is given the same weight.

Table 8. Comparison of the proposed ensemble framework with several standard convolution neural network models in the literature on both the Kermany and the Radiological Society ofNorth America challenge datasets.

Dataset	Model	Acc(%)	Pre(%)	Rec(%)	F1(%)	AUC(%)
Kermany	GoogLeNet	97.89	98.12	98.12	98.12	97.89
	AlexNet	97.17	97.22	97.18	97.19	97.17
	VGG-16	97.09	97.12	97.09	97.1	97.09
	DenseNet-121	96.23	96.63	96.24	96.31	96.23
	ResNet-18	97.29	98.31	98.29	98.3	97.29
	Proposed Method	**98.81**	**98.82**	**98.80**	**98.79**	**98.35**
RSNA	GoogLeNet	84.83	84.98	84.83	84.90	84.83
	AlexNet	85.86	85.15	84.86	85.00	85.86
	VGG-16	81.05	80.98	81.17	80.62	81.05
	DenseNet-121	85.67	84.98	85.35	85.26	85.67
	ResNet-18	84.98	85.35	85.37	85.46	84.98
	Proposed Method	**86.85**	**86.89**	**87.02**	**86.95**	**86.85**

The same base learners were used in all the ensembles: GoogLeNet, ResNet-18, and DenseNet-121.

9. CONCLUSION AND FUTURE WORK

To treat pneumonia appropriately and keep the patient's life from being in danger, early recognition of the illness is essential. The most common method for diagnosing pneumonia is a chest radiograph; however, there can be inter-class variation in these images, and the diagnosis relies on the doctor's skill to identify early signs of pneumonia. In this work, an automated

CAD system was created to aid medical professionals. It classifies chest X-ray pictures into two groups, "Normal" and "Pneumonia," using deep transfer learning-based classification. A weighted average ensemble is formed by an ensemble framework that takes into account the decision scores from three CNN models: GoogLeNet, ResNet-18, and DenseNet-121. An innovative technique was used to calculate the weights assigned to the classifiers, whereby five evaluation parameters, accuracy, precision, recall, f1-score, and AUC, were fused using the hyperbolic tangent function. The framework, evaluated on two publicly available pneumonia chest X-ray datasets, obtained an accuracy rate of 98.2%, a sensitivity rate of 98.19%, a precision rate of 98.22%, and an f1-score of 98.29% on the Kermany dataset and an accuracy rate of 86.7%, a sensitivity rate of 86.62%, a precision rate of 86.69%, and an f1-score of 86.65% on the RSNA challenge dataset, using a five-fold cross-validation scheme. It outperformed state-of-the-art methods on these two datasets. Statistical analyses of the proposed model using McNemar's and ANOVA tests indicate the viability of the approach. Furthermore, the proposed ensemble model is domain-independent and thus can be applied to a large variety of computer vision tasks.

But as was previously mentioned, there were instances in which the ensemble architecture was unable to produce precise estimates. To improve the quality of the photos, I could investigate techniques like picture contrast enhancement or other pre-processing steps in the future. Before categorizing the lung picture, I recommend segmenting it to assist the CNN models extract additional characteristics from it. Furthermore, since three CNN models are required to train the recommended ensemble, its computational cost is higher than that of the CNN baselines developed in works published in the literature. To try to reduce the processing needs in the future, I may consider employing strategies like snapshot ensembling increasing overall performance.

ACKNOWLEDGEMENTS

I would like to acknowledge my parents for buying me the computer on which I did all of my research.

REFERENCES

[1] WHO Pneumonia. World Health Organization. (2019), https://www.who.int/news-room/fact-sheets/detail/pneumonia

[2] Kermany D., Zhang K. & Goldbaum M. Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification. (Mendeley,2018)

[3] Dalhoumi S., Dray G., Montmain J., Derosière, G. & Perrey S. An adaptive accuracy-weighted ensemble for inter-subjects classification in brain-computer interfacing. 2015 7th International IEEE/EMBS Conference On Neural Engineering (NER). pp. 126-129 (2015)

[4] Albahli S., Rauf H., Algosaibi A. & Balas V. AI-driven deep CNN approach for multi-label pathology classification using chest X-Rays. PeerJ Computer Science. 7 pp. e495 (2021) pmid:33977135

[5] Rahman T., Chowdhury M., Khandakar A., Islam K., Islam K., Mahbub Z., et al. Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray. Applied Sciences. 10, 3233 (2020)

[6] Liang G. & Zheng L. A transfer learning method with deep residual network for pediatric pneumonia diagnosis. Computer Methods And Programs In Biomedicine. 187 pp. 104964 (2020) pmid:31262537

[7] Ibrahim A., Ozsoz M., Serte S., Al-Turjman F. & Yakoi P. Pneumonia classification using deep learning from chest X-ray images during COVID-19. Cognitive Computation. pp. 1–13 (2021) pmid:33425044

[8] Zubair S. An Efficient Method to Predict Pneumonia from Chest X-Rays Using Deep Learning Approach. The Importance Of Health Informatics In Public Health During A Pandemic. 272 pp. 457 (2020)

[9] Rajpurkar P., Irvin J., Zhu K., Yang B., Mehta H., Duan T., et al. & Others Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. ArXiv Preprint ArXiv:1711.05225. (2017)

[10] Albahli S., Rauf H., Arif M., Nafis M. & Algosaibi A. Identification of thoracic diseases by exploiting deep neural networks. Neural Networks. 5 pp. 6 (2021)

[11] Chandra T. & Verma K. Pneumonia detection on chest X-Ray using machine learning paradigm. Proceedings Of 3rd International Conference On Computer Vision And Image Processing. pp. 21-33 (2020)

[12] Kuo K., Talley P., Huang C. & Cheng L. Predicting hospital-acquired pneumonia among schizophrenic patients: a machine learning approach. BMC Medical Informatics And Decision Making. 19, 1–8 (2019) pmid:30866913

[13] [13] Yue H., Yu Q., Liu C., Huang Y., Jiang Z., Shao C., et al. & Others Machine learning-based CT radiomics method for predicting hospital stay in patients with pneumonia associated with SARS-CoV-2 infection: a multicenter study. Annals Of Translational Medicine. 8 (2020) pmid:32793703

[14] Sharma H., Jain J., Bansal P. & Gupta S. Feature extraction and classification of chest x-ray images using cnn to detect pneumonia. 2020 10th International Conference On Cloud Computing, Data Science & Engineering (Confluence). pp. 227-231 (2020)

[15] Stephen O., Sain M., Maduh U. & Jeong D. An efficient deep learning approach to pneumonia classification in healthcare. Journal Of Healthcare Engineering. 2019 (2019) pmid:31049186

[16] Wang X., Peng Y., Lu L., Lu Z., Bagheri M. & Summers R. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localTocommon thorax diseases. Proceedings Of The IEEE Conference On Computer Vision And Pattern Recognition. pp. 2097-2106 (2017)

[17] Selvaraju R., Cogswell M., Das A., Vedantam R., Parikh D. & Batra D. Grad-cam: Visual explanations from deep networks via gradient-based localization. Proceedings Of The IEEE International Conference On Computer Vision. pp. 618-626 (2017)

[18] Mahmud T., Rahman M. & Fattah S. CovXNet: A multi-dilation convolutional neural network for automatic COVID-19 and other pneumonia detection from chest X-ray images with transferable multi-receptive feature optimization. Computers In Biology And Medicine. 122 pp. 103869 (2020) pmid:32658740

[19] Dietterich T. Approximate statistical tests for comparing supervised classification learning algorithms. Neural Computation. 10, 1895–1923 (1998) pmid:9744903

[20] Cuevas A., Febrero M. & Fraiman R. An anova test for functional data. Computational Statistics & Data Analysis. 47, 111–122 (2004)

AUTHOR

Prithvi is a driven high school student set to graduate in 2025 with an impressive academic record in computer science and STEM fields. His diverse pursuits, ranging from founding a global AI education platform to pioneering research in AI-powered medical imaging analysis, exemplify his innovative mindset and passion for developing socially impactful technology. With an unwavering determination to be at the forefront of ethical AI innovation, Prithvi aims to continue pushing boundaries and creating lasting positive impacts in the field.