Pneumonia Detection using Deep Learning and Ensemble Learning

Cristian Agusta
Computer Science Department, School of Computer Science
Bina Nusantara University
Jakarta, Indonesia
cristian.agusta@binus.ac.id

Makaio Kanaka
Computer Science Department, School of Computer Science
Bina Nusantara University
Jakarta, Indonesia
makaio.kanaka@binus.ac.id

Dimas Ramdhan
Computer Science Department, School of Computer Science
Bina Nusantara University
Jakarta, Indonesia
dimas.ramdhan@binus.ac.id

David David
Computer Science Department, School of Computer Science
Bina Nusantara University
Jakarta, Indonesia
david01@binus.edu

Abstract— **Pneumonia is one of the leading causes of death in children under five years old. Detecting pneumonia requires an expert examining a patient's chest X-ray. This process takes a lot of time and manpower. To remedy this, there has been research done to create an Artificial Intelligence that could detect and diagnose pneumonia.**

To further this research, we propose an Ensemble Deep Learning Model composed of VGG-16, Xception, and EfficientNetB7. We have found that the Ensemble model with the highest accuracy is the one composed of the Xception and EfficientNetB7 models with a score of 84%. However, this accuracy is still below that of the VGG-16 model.

Keywords— Pneumonia, Deep Learning, Ensemble Learning, Convolutional Neural Network, Classification

I. INTRODUCTION

Pneumonia is one of many potentially life-threatening diseases. Throughout the world, pneumonia is a leading cause of death for children younger than 5 years. Pneumonia is an inflammatory condition of the lung, causing pus or fluids to fill it, caused by pathogens such as viruses and bacteria. Pneumonia itself is generally divided into 2, namely Bacterial Pneumonia which is caused by the entry of Bacterial microorganisms and Viral Pneumonia which is caused by exposure to Viruses.

Pneumonia is one of the most difficult diseases to detect, because it occurs in the inner part of the human body, especially the lungs. It takes complex analysis and diagnosis of the patient’s chest X-Ray before a conclusion can be drawn whether there is pneumonia or not. The problem arises because sometimes health services lack an expert who can analyze the X-Ray. Therefore, a method that allows for effective and comprehensive pneumonia detection is needed.

Rapid technological advancements play a significant role in diagnosing pneumonia. The existence of a deep learning method, which allows a computer to work and think like a human, facilitates the analysis process. Deep Learning models, such as CNN (Convolutional Neural Network), learn and extract each feature from a dataset. For example, chest X-Rays images in cases like Pneumonia. Each dataset will be trained by connecting it to the corresponding class, to find the general characteristics of each class. These characteristics will later be used by the model to classify Pneumonia.

There is a constant need to improve on Deep Learning Models to better diagnose pneumonia. One such method is to combine preexisting models into an ensemble deep learning model. By combining models, one can achieve higher diagnosis speed and accuracy than by simply using a single model. This paper will test the performance of various deep learning models in diagnosing pneumonia and comparing their performance with that of an ensemble model.

II. Literature Review

Application of Deep Learning in Detecting Pneumonia, in which using Computer as foundation, meaning that there are several ways to increase the accuracy of prediction made by proposed model. Reshan, M. et al. are comparing several architectures of pre-Trained Models, within a stack of parameter to find the best and most optimal model, finding that MobileNet has the most optimal model compared to others with the given dataset with the accuracy of 94.23%[1]. Jaiswal, A.K. et al. explained the Object Detection method called Mask-RCNN, which is a level of Faster-RCNN. Mask-RCNN is used to classify objects and perform segmentations on images. Mask-RCNN is built with ResNet architecture (ResNet-50 and ResNet-101) as the basis for Feature Extraction. Finally, an Ensemble Model is formed from ResNet-50 and ResNet-101 to provide more accurate results in determining Pneumonia[2]. Zhang, D. et al. proposes a method where the image is being augmented by DHE, rather than only using original image, which further increases the accuracy of the model to 96.07%, higher than other models [3]. Hu, Z. et al. took a different approach on augmenting the image datasets. In this paper, the authors are using a method called Radiomics Map Feature which brings a Saliency Map into every image dataset. This method increases the accuracy of model to 97% compared to standard CNN without using Radiomics Map Feature which has an accuracy of 93%[4].

In addition to analyzing pneumonia disease, Deep Learning, especially CNN models, are widely used to analyze many diseases, such as Diabetic Retinopathy and COVID-19. Jena, P. K. et al. in their paper use the CLAHE method to augment retinal images using the Deep Learning U-Net method to classify the severity of Diabetic Retinopathy. U-Net itself is a model used to perform segmentation on images related to biomedicine such as blood vessels, to obtain more meaningful information. Then, after the image is augmented, the trained CNN model is used and the classification process is done using SVM [5].

Mercaldo, F. et al. in their paper used the VGG-16 architecture to analyze the characteristics of the lungs of COVID-19 patients. The Grad-CAM approach is also used to capture the most prominent and most determining visual features in providing the model output. The model created with the VGG-16 architecture and the Grad-CAM approach can provide an accuracy of 95%, which is higher than ResNet-50 (92%), VGG19 (93%), AlexNet (71%), and InceptionV3 (84%) [[6]. Atallah, O. et al. proposed a different approach regarding COVID-19 Detection. In this paper, the author suggested DWT (Discrete Wavelet Decomposition) to further enhance the accuracy of the proposed model. The author also uses SVM as classifier which gives the model a final accuracy of 99.23% and 99.62% for the OMNIAHCOV and SARS-COV-2-CT-Scan datasets, respectively, which are superior to related studies [[7]. Meanwhile in this paper, the author is using not only DWT method, but also GLCM method in process, which gives an accuracy of 99.60% for the given dataset [[8]. Li, X. et al. proposes a method called Ensemble Learning, in which combining the result taken from VGG-16 Architecture and Stacked Generalization Ensemble Learning, getting an accuracy of 93.57%[9]. Ahuja, S. et al. ara proposing a 3-phase method, Data Augmentation phase, Model Training phase, and Abnormality Localization phase, using 4 different Architecture, giving ResNet18 the best accuracy of 99.82% for the given method and the given dataset[10].

III. METHODOLOGY

3.1 Datasets

The Chest X-Ray image dataset consists of two classes, healthy Chest X-Ray images and Pneumonia Chest X-Ray images. The dataset is downloaded from https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia. This dataset has a total of 5863 images that is divided into 3 folders in which train dataset, test dataset, and validation dataset.

Chest X-ray images (anterior-posterior) were selected from retrospective cohorts of pediatric patients of one to five years old from Guangzhou Women and Children's Medical Center, Guangzhou. All chest X-ray imaging was performed as part of patients' routine clinical care.

For the analysis of chest x-ray images, all chest radiographs were initially screened for quality control by removing all low quality or unreadable scans. The diagnoses for the images were then graded by two expert physicians before being cleared for training the AI system. In order to account for any grading errors, the evaluation set was also checked by a third expert.

3.2 Data Augmentation

Before Datasets being trained into models proposed, augmentation is being done to models to increase variety and variance of datasets. Increasing variety and variance of datasets is done to make sure models could learn lots of patterns from different images. There are a lot of steps in augmenting image-based datasets, such as, shifting, shearing, rotating, zooming, and flipping.

First, the dataset is rotated 40 degrees clockwise or counterclockwise. Other than being rotated, images are shifted by up to 20% horizontally and vertically. The images are also zoomed in or out by up to 20% and sheared by up to 20%. All these image augmentations are done randomly. Randomly means that every dataset is being augmented randomly from 0 to the parameter value of every parameter.

3.3 Ensemble Learning

Ensemble Learning is a method in Machine Learning proposed to combine some models to bring a different value to prediction being made.

The ensemble learning model will be using a combination voting method. Multiple models will be used to diagnose the same Chest X-Ray images. The diagnosis from each model will be used as a "vote" and the final answer taken is majority diagnosis.

3.4 Proposed Model

We will be testing the performance of ensemble models composed of three CNN models: VGG-16, Xception, and EfiicientNetB7. CNN Model consists of a few layer as a way to make a prediction, as can be seen from Fig. 1. We will test each model's performance by themselves. After we tested each of the model, we will compare each and every model to see which model actually have better accuracy on given dataset singularly. Every model will be trained using the same number of epochs, which is 5 and validation step used is 8.

After doing it singularly, model will be combined using Ensemble Learning into four different combinations. First combinations will be constucted by VGG16 model with Xception model. The second one will be constructed between VGG16 and EfficientNetB7 model. While the last one will be constructed by the combination of Xception Model and EfficientNetB7 model. After these ensemble models being constructed, they will be compared each other and other singular model to see whether Ensemble Learning will bring an increasing of accuracy for models proposed.

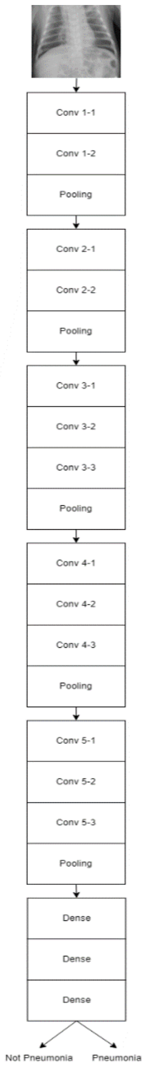

Fig 1. Architecture of CNN Model

IV. EXPERIMENTAL RESULTS

4.1. Individual Model Results

Each model is trained with the same hyperparameters, namely the number of epochs is 5 and validation steps are 8. In the previous experiment, with 10 epochs the validation_loss value received was very high causing the model to have low accuracy. As can be seen from Table 1, when the models were trained individually, the VGG16 model had a fairly high level of accuracy, namely 86.9%. Other models such as EfficientNetB7 have an accuracy of 83.23% and Xception has an accuracy of 80.51%.

Model	**Accuracy Score**
VGG-16	86.90%
Xception	80.51%
EfficientNetB7	83.23%

Table 1 The Accuracy Result of each Individual Model

4.2. Ensemble Model Performance

According to Table 2, when the model was trained using ensemble learning, Ensemble Model 1, namely VGG16 and Xception, had an accuracy value of 77.35%, which means the accuracy value was lower than the VGG16 and Xception models individually. The same thing also happened to the Ensemble 2 models, namely VGG16 and EfficientNetB7, which only had an accuracy of 79.43%, lower than when the models were trained individually. Improvements occurred for the Ensemble 3 model consisting of Xception and EfficientNetB7. The Ensemble 3 model produces an accuracy of 84.23%, which is higher than the Xception and EfficientNetB7 models which were trained individually, namely 80.51% and 83.23%.

Ensemble Model	**Composistion**	**Accuracy Score**
Ensemble Model 1	VGG-16 and Xception	77.35%
Ensemble Model 2	VGG-16 and EfficientNetB7	79.43%
Ensemble Model 3	Xception and EfficientNetB7	84.23%

Table 2 The Accuracy of Each Ensemble Model

Based on experiment that has been done, it can be seen that each classification model will provide different results. This is because there are differences in the architecture of each model which results in the data training process carried out by each model being different, thus providing different prediction results.

Among the ensemble models, the only ensemble model that improved the accuracy of its composing models is the one composed of Xception and EfficientNetB7. Despite being partially composed of VGG16, the model with the highest individual accuracy score, other ensemble models have a lower accuracy score than their composing models.

V. CONCLUSION

It can be concluded that Ensemble Learning, like other Deep Learning models, has certain compatibility. Using Ensemble Learning does not necessarily mean that the combined model will provide higher prediction results and accuracy. Many experiments are needed in tuning each model and seeing the suitability of the model in making predictions to ensure that the learning ensemble formed provides satisfactory results.

REFERENCES

[1] M. S. Al Reshan *et al.*, "Detection of Pneumonia from Chest X-ray Images Utilizing MobileNet Model," *Healthcare (Switzerland)*, vol. 11, no. 11, Jun. 2023, doi: 10.3390/healthcare11111561.

[2] A. K. Jaiswal, P. Tiwari, S. Kumar, D. Gupta, A. Khanna, and J. J. P. C. Rodrigues, "Identifying pneumonia in chest X-rays: A deep learning approach," *Measurement (Lond)*, vol. 145, pp. 511–518, Oct. 2019, doi: 10.1016/j.measurement.2019.05.076.

[3] D. Zhang, F. Ren, Y. Li, L. Na, and Y. Ma, "Pneumonia detection from chest x-ray images based on convolutional neural network," *Electronics (Switzerland)*, vol. 10, no. 13, Jul. 2021, doi: 10.3390/electronics10131512.

[4] Z. Hu, Z. Yang, K. J. Lafata, F.-F. Yin, and C. Wang, "A Radiomics-Boosted Deep-Learning Model for COVID-19 and Non-COVID-19 Pneumonia Classification Using Chest X-ray Image."

[5] P. K. Jena, B. Khuntia, C. Palai, M. Nayak, T. K. Mishra, and S. N. Mohanty, "A Novel Approach for Diabetic Retinopathy Screening Using Asymmetric Deep Learning Features," *Big Data and Cognitive Computing*, vol. 7, no. 1, Mar. 2023, doi: 10.3390/bdcc7010025.

[6] F. Mercaldo, M. P. Belfiore, A. Reginelli, L. Brunese, and A. Santone, "Coronavirus covid-19 detection by means of explainable deep learning," *Sci Rep*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-27697-y.

[7] O. Attallah and A. Samir, "A wavelet-based deep learning pipeline for efficient COVID-19 diagnosis via CT slices," *Appl Soft Comput*, vol. 128, Oct. 2022, doi: 10.1016/j.asoc.2022.109401.

[8] O. Attallah, "A computer-aided diagnostic framework for coronavirus diagnosis using texture-based radiomics images," *Digit Health*, vol. 8, Apr. 2022, doi: 10.1177/20552076221092543.

[9] X. Li, W. Tan, P. Liu, Q. Zhou, and J. Yang, "Classification of COVID-19 Chest CT Images Based on Ensemble Deep Learning," *J Healthc Eng*, vol. 2021, 2021, doi: 10.1155/2021/5528441.

[10] S. Ahuja, B. K. Panigrahi, N. Dey, V. Rajinikanth, and T. K. Gandhi, "Deep transfer learning-based automated detection of COVID-19 from lung CT scan slices," *Applied Intelligence*, vol. 51, no. 1, pp. 571–585, Jan. 2021, doi: 10.1007/s10489-020-01826-w.