ISSN 2442-6571

Pneumonia detection on x-ray imaging using softmax output in multilevel meta ensemble algorithm of deep convolutional neural network transfer learning models

Simeon Yuda Prasetyo [a,1,*], Ghinaa Zain Nabiilah [a,2], Zahra Nabila Izdihar [a,3], Sani Muhamad Isa [b,4]

[a] Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta 11480, Indonesia
[b] Computer Science Department, BINUS Graduate Program - Master of Computer Science, Jakarta and 11530, Indonesia
[1] simeon.prasetyo@binus.ac.id; [2] ghinaa.nabiilah@binus.ac.id; [3] zahra.izdihar@binus.ac.id; [4] sani.m.isa@binus.ac.id
* corresponding author

ARTICLE INFO

Article history
Received August 02, 2022
Revised April 2, 2023
Accepted April 16, 2023
Available online July 8, 2023

Keywords
Pneumonia classification
Transfer learning
Deep convolutional neural network
Softmax
Multilevel meta ensemble

ABSTRACT

Pneumonia is the leading cause of death from a single infection worldwide in children. A proven clinical method for diagnosing pneumonia is through a chest X-ray. However, the resulting X-ray images often need clarification, resulting in subjective judgments. In addition, the process of diagnosis requires a longer time. One technique can be applied by applying advanced deep learning, namely, Transfer Learning with Deep Convolutional Neural Network (Deep CNN) and modified Multilevel Meta Ensemble Learning using Softmax. The purpose of this research was to improve the accuracy of the pneumonia classification model. This study proposes a classification model with a meta-ensemble approach using five classification algorithms: Xception, Resnet 15V2, InceptionV3, VGG16, and VGG19. The ensemble stage used two different concepts, where the first level ensemble combined the output of the Xception, ResNet15V2, and InceptionV3 algorithms. Then the output from the first ensemble level is reused for the following learning process, combined with the output from other algorithms, namely VGG16 and VGG19. This process is called ensemble level two. The classification algorithm used at this stage is the same as the previous stage, using KNN as a classification model. Based on experiments, the model proposed in this study has better accuracy than the others, with a test accuracy value of 98.272%. The benefit of this research could help doctors as a recommendation tool to make more accurate and timely diagnoses, thus speeding up the treatment process and reducing the risk of complications.

1. Introduction

Pneumonia is an inflammation of the lungs due to an acute respiratory infection caused by viruses, bacteria, or fungi. In a healthy state of breathing, the lungs will consist of tiny sacs (alveoli) filled with air. In contrast, the alveoli will be filled with pus and fluid for people with pneumonia, making breathing painful and limiting oxygen intake. Pneumonia is the leading cause of death due to children's single most extensive infection worldwide. In 2019, around 740,180 children under five died [1]. While data in Indonesia in 2019 revealed that about 314,455 children under five died due to pneumonia based on data from The Ministry of Health of the Republic of Indonesia [2].

Pneumonia can be treated using medications such as antibiotics as the primary step to preventing death due to complications of pneumonia. Early diagnosis can be made so people with pneumonia can get early treatment. A well-known and effective clinical method for diagnosing pneumonia is chest X-ray images [3]. However, diagnosing pneumonia from chest X-ray images is still incredibly challenging, even for radiologists. X-ray images are often not clear, so it can confuse the results of a diagnosis with other diseases [4]. This results in inconsistencies leading to many subjective decisions, and there are differences between radiologists in diagnosing pneumonia. In addition, radiologists' analysis and diagnosis process often requires a long time because it is still done manually.

A computer-based system is necessary to aid radiologists in detecting pneumonia through analyzing chest X-ray images. The development of machine learning models has good ability in the classification process, including image classification. Chandra and Verma [5] have proven the ability of machine learning models to detect chest X-ray images by comparing several machine learning classification models such as Regression Logistics, Multilayer Perceptron, Random Forest, and Sequential Minimal Optimization. This study also used feature extraction before classification. The optimal result of this research is using logistic regression with 95.53% accuracy. Although it is accurate, machine learning models have shortcomings in managing large amounts of data. The optimal results obtained from the study [6] sed a limited amount of data with 412 image data from chest X-rays.

The development research related to image classification is more developed using deep learning methods. Deep learning methods are considered to have an excellent ability to handle large amounts of data. One of the best methods for image classification is to use the Convolutional Neural Network (CNN) along with various variations of CNN [7]. CNN is also successfully used in classifying images for medical problems, such as detecting breast cancer, lung cancer, brain tumor segmentation, and skin diseases [8], [9].

The idea of Transfer Learning, which involves acquiring new knowledge by leveraging existing knowledge, was introduced by Pan *et al.* in 2010. Essentially, this approach involves identifying commonalities between known and unknown knowledge. Since some knowledge areas may complex to learn from scratch, the use of existing knowledge to facilitate learning is crucial [10].

The fundamental concept of Transfer Learning is to find the correlation between familiar and unfamiliar knowledge to obtain new insights. Transfer Learning refers to the process of transferring knowledge from a known domain (called Source Domain) to an unknown domain (called Target Domain). The primary aim of Transfer Learning is to explore ways to transfer knowledge from the Source Domain to the Target Domain. In the field of machine learning, Transfer Learning aims to use existing models to apply previously acquired knowledge to new knowledge. There are four categories of learning approaches used in Transfer Learning based on their learning methods, which include features, model-based, relationship, and sample-based methods [11].

Research on image classification for pneumonia using deep learning has previously been carried out. As research conducted by Enes and Murat [12] uses two modified models, namely Xception and a Vgg 16-based model that uses transfer learning, fine-tuning, and data augmentation in Categorizing medical images of chest X-rays for individuals with pneumonia. This study showed that the Xception model outperformed the Vgg 16 model in detecting pneumonia. In 2021, Zhang *et al.* [13] modified the architecture of the VGG model by using layers and with fewer parameters. The results of this study compared with other model architectures such as VGG-16, RES-50, Xception, DenseNet21, and MobileNet. However, the performance of the proposed model is superior to those models with 96.068% of accuracy and 0.99107 AUC. Based on this research, in this study, an experiment was carried out by applying ensemble stacking modifications from several model architectures such as Xception, Resnet152V2, InceptionV3, VGG16, and VGG19. This modification was later named Multilevel Ensemble Stacking. With the modification experiment in this study, it is hoped that the results of the diagnostic image from X-rays of pneumonia patients can be more accurate.

Several techniques have been presented in recent years to illustrate the brief process of detecting pneumonia based on chest X-ray images using deep learning or deep CNN methods. The deep CNN method is more widely used and is considered to be successfully applied to improve computer performance, especially in image classification in the medical field. An example of implementation is the research conducted by Radjpukar [14] which uses a CNN model with 121 layers to detect pneumonia using a chest X-ray image. The dataset used in this study has more than 100,000 images with 14 other diseases associated with chest X-rays, including pneumonia. This study had different results, with radiologists having higher F1 scores detecting all 14 diseases, including pneumonia. Rahman *et al.* conducted a study [15] that utilized a transfer learning approach based on Deep CNN to detect pneumonia. Four deep learning algorithms based on CNN, namely AlexNet, ResNet18, DenseNet201, and SqueezeNet, were trained to classify chest X-ray images of pneumonia and non-pneumonia patients. The results showed that DenseNet201 outperformed the other three deep CNN networks, achieving an optimal accuracy of 98%.

The success of deep learning methods in analyzing images makes Deep CNN models get a lot of attention for classifying images. However, research conducted by Varshni *et al.* [16] tried to use machine learning methods such as SVM, Naive Bayes, KNN, and Random Forest, which is also well known in the classification process as a method for classifying images. Deep CNN models such as XCeption, VGG-16, VGG-19, ResNet-50, DenseNet-121, and DenseNet-169 were also used in this study as feature extraction. The optimal result of this research is to use DenseNet-169 and SVM with AUC 0.8002.

Due to the success of the deep CNN model in classifying images, the latest research related to image classification tries to add the application of ensemble learning to get better accuracy results by combining some of the best models. Research conducted by Chouhan *et al.* [17] implements ensemble learning that combines Deep CNN models such as AlexNet, DeenseNet121, InceptionV3, ResNet18, and GoogleNet. The results of this study reached an accuracy of 96.4%. Research that applies ensemble learning has also been carried out by Mubrouk *et al.* [18] which combines Deep CNN models such as Vision Transformer, MobileNetV2, and DenseNet169. The optimal result of this research is an accuracy value of 93.91%.

Deep CNN is an alternative learning method because it can identify patterns automatically, even in low-resolution images. Junior *et al.* [19] proposed a deep CNN model with different training strategies. This study evaluates the preprocessing of different images to improve classification by image cropping and histogram equalization. In this research, the ensemble technique uses the VGG16 architecture by modifying the fully connected layer with the Multilayer Perceptron.

The goal of this research is to create and assess a system that can identify pneumonia in chest X-ray images by implementing a multilevel meta-ensemble method. This approach involves combining the predictions of multiple models at different levels of abstraction, using the Softmax activation function as the output. The motivation for using the multilevel meta-ensemble method is to improve the model's performance in detecting pneumonia in chest X-ray images. In the detection of pneumonia in chest X-ray images, there are often variations in image quality, such as lighting, contrast, and radiograph quality. Therefore, by using the multilevel meta-ensemble method, it is expected that the model can learn and extract output features (in the form of Softmax or probabilities output) from various high-level feature maps features, thereby improving the accuracy of pneumonia detection.

The contribution of using the multilevel meta ensemble method is to improve accuracy in detecting pneumonia in chest X-ray images. By utilizing this technique, the model can leverage information derived from various output probabilities (Softmax output), thereby enhancing its ability to distinguish between pneumonia (inflammation) and normal conditions in X-ray images. Furthermore, the use of multilevel meta ensemble method can also contribute to the development of more advanced medical technology. By improving the accuracy of detecting pneumonia in X-ray images, this technology can assist doctors in making more accurate and timely diagnoses, thereby speeding up the treatment process and reducing the risk of complications.

Overall, the use of multilevel meta ensemble method in detecting pneumonia in X-ray images has a strong motivation and contribution to improving the accuracy of detecting this disease, as well as providing a positive impact on the development of medical technology as a whole.

Vol. 9, No. 2, July 2023, pp. 319-330

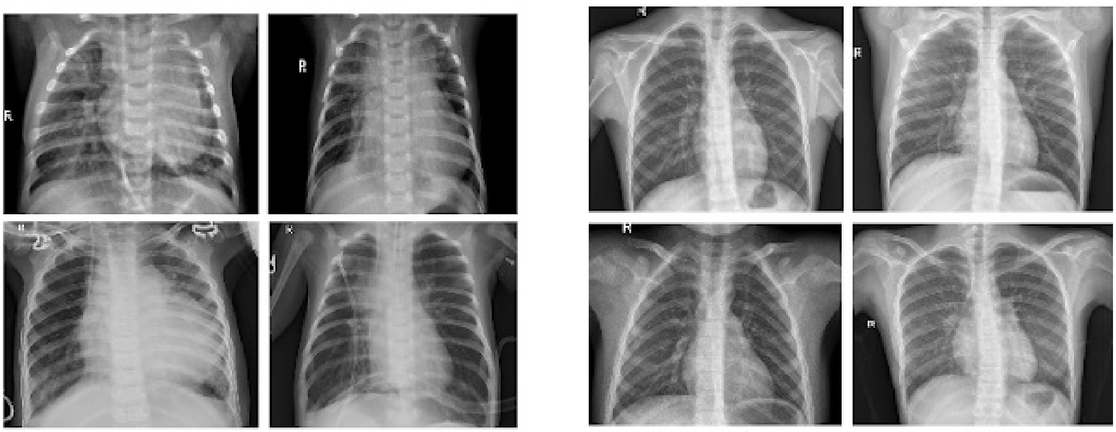
2. Method

2.1. Dataset Collection

This study used an open-source dataset created by Mooney and sourced from Kaggle [20]. The dataset contains approximately 5,865 Chest X-Ray Images, which are categorized into two groups: Pneumonia and Normal. The chest X-ray images used in this study were chosen from a group of previous cases of pediatric patients between the ages of one and five who had undergone chest X-ray imaging during their regular clinical checkups at the Guangzhou Women and Children's Medical Center in Guangzhou. These images were obtained using the anterior-posterior view, which is the usual technique for capturing this type of image [21].

The normal chest X-ray, which is shown in the left panel, depicts lungs that are clear without any visible areas of abnormal opacification. Bacterial pneumonia, as depicted in the middle panel, typically shows a focal lobar consolidation, which means that there is a localized area of the lung appearing dense on the X-ray image. In this case, the consolidation is in the right upper lobe and can be identified by the white arrows. On the other hand, viral pneumonia, as shown in the right panel, has a more diffuse "interstitial" pattern that appears throughout both lungs. This pattern is characterized by a fine, reticular shadowing that is visible on the X-ray image. Understanding these visual characteristics of pneumonia on chest X-rays is important for accurate diagnosis and treatment of patients.

In order to ensure that the data used in the study is reliable, the chest x-ray images were subjected to quality control screening, and any images of low quality or that were unreadable were excluded. The diagnoses for the remaining images were then assessed by two expert physicians to guarantee the accuracy of the training data used in the AI system. To minimize the potential impact of grading errors, a third expert also evaluated the dataset. The Pneumonia dataset has data distribution which is divided into training data, test data and validation with 4695, 521 and 624, respectively. The pneumonia dataset consists of various images of chest X-rays of pneumonia patients and non-pneumonia patients. Fig. 1 contain an example of a dataset image.



(a) Sample Data of Pneumonia Total Data: 1.583 images

(b) Sample Data of Non- Pneumonia Total: 4.237 images

Fig. 1. Example of a dataset image

2.2. Data Preprocesing

In order to standardize the dataset and account for discrepancies in the size of the X-ray images, a decision was made to resize all images to a uniform size of 224 x 224 pixels. Since the dataset is not consistent and images may vary in size, this process is crucial to ensure accurate and reliable analysis. To accomplish this, RGB reordering has been implemented, resulting in a final input of 224 × 224 × 3 for the proposed model. This standardization of the images enables the use of established models and facilitates the analysis process.

The use of a uniform size for all images eliminates inconsistencies that may arise from differences in image sizes, ensuring the accuracy of the analysis. Given that the data set was not uniform and comprised of images of different sizes, it was necessary to transform all images to 224 x 224 pixels. RGB reordering was employed to make the images consistent and provide a final input of 224 × 224 × 3 to the proposed model. This standardization of the images ensures that the model can effectively process the data set, improving the accuracy and reliability of the analysis.

2.3. Classification using Deep Convolutional Neural Network

The convolutional neural network (CNN) is a type of neural network that is proficient in handling large amounts of data with high precision and minimal computational expense due to its multiple layers. The fundamental architecture of CNN includes various layers, such as convolutional, flattening, pooling and dense layers [22].

After conducting initial preprocessing, the following step is to extract features that meet the criteria. The suggested approach in this study involves utilizing a multilevel ensemble stacking method in a transfer learning model of a Convolutional Neural Network (CNN). CNN offers numerous benefits, including the ability to handle both feature extraction and classification tasks with a single structure. Additionally, this network can extract more complex 2-D features and is completely adaptive, ensuring invariance to geometric and local changes in the image [23].

The Convolutional Neural Network (CNN) consists of three main types of layers: Convolution layer, Pooling layer (Subsampling), and Output layer. These layers are arranged in a feed-forward structure within the network. The convolution layer is always followed by a pooling layer, while the last convolution layer is followed by an output layer. The convolution and pooling layers are 2-D layers, whereas the output layer is a 1-D layer. Each 2-D layer of the CNN comprises multiple planes, where each plane is made up of a 2-D array of neurons. The output of a plane is known as a feature map. The proposed methodology's architectures will be explained in detail in the following section [24].

In this research, several pretrained models such as InceptionV3, Xception, VGG, and ResNet will be used, which will then be combined using a multilevel meta ensemble approach that utilizes output in the form of probabilities (using Softmax activation function). Inception is a development model of the first CNN [25]. Unlike the Inception-V3 which is a combination of all improvements from Inception V2 which uses Label Smoothing as a component to prevent overfitting, factorized 7 x 7 convolutions, and additional classifier equalization to bring label information to the lower network accompanied by normalization. batch on the layer on the side head [26].

Xception (extreme inception) is a convolutional neural network with 71 deep layers. The Xception model architecture is a convolution that can be separated in depth by improving the Inception model architecture [27]. In the Xception model, a revolutionary deep convolutional neural network architecture is proposed. This causes the Xception model to outperform the V3 inception in the case of image classification with larger data, consisting of 350 million images and 17,000 classes. More effective use of model parameters rather than additional capacity makes for increased performance on the Xception model. The parameters used in the Xception model are also the same as those in the InceptionV3 model [28].

VGG is an architectural model consisting of 16 trained layers that have weights. The VGG model architecture contains a CNN model architecture that utilizes a convolutional layer specification of a small 3 x 3 convolutional filter. Due to the small size of the convolutional filter, the neural network can have more convolutional layers [29]. This model was created in 2015 and trained on one million images from the ImageNet database [30].

Residual Networks or ResNets were created to solve the gradient problem. The ResNet architecture introduces the concept of residual block or skip connections. The activation layer connects to the next layer by passing through several layers to form a residual block. To make a network, stack leftover blocks on top of each other; for example, ResNet-50 uses this block in fifty layers [31].

The selected configuration for the pretrained model involves utilizing the weight architecture solely for feature extraction in the convolutional section. This means that during the training process, the weights are adjusted to help the model learn to identify important features in the input data. Once the high-level feature maps are obtained from the convolution, they are flattened, or reshaped into a single vector, and then directly fed into a fully connected layer with a size of 2. In this layer, each of the flattened features is connected to every neuron in the layer. Finally, Softmax activation function is applied to the outputs of the fully connected layer, which allows the model to output a probability distribution over the possible classes that the input data could belong to.

2.4. Ensemble Method

Ensemble learning is how an algorithm learns data using a combination of several algorithms or models (where several learning algorithms are used simultaneously, then combined) to get output with higher accuracy when compared to using only one algorithm. Several types of Ensemble learning include Bagging, Boosting, and Stacking [32].

In this study, a classification system with a meta ensemble approach is proposed to detect chest X-rays of patients with pneumonia and non-pneumonia patients. The classifier combines the predictions of several models to conclude the final result [33]. Specifically, in this study, there are five different classification models on the data that have been processed previously and then assigned a weighting factor to the predictions of each model. The meta-ensemble learning model aims to fit complex data better, lowering uncertainty estimates. Fig. 2 contains an illustration of the usual meta ensemble process.

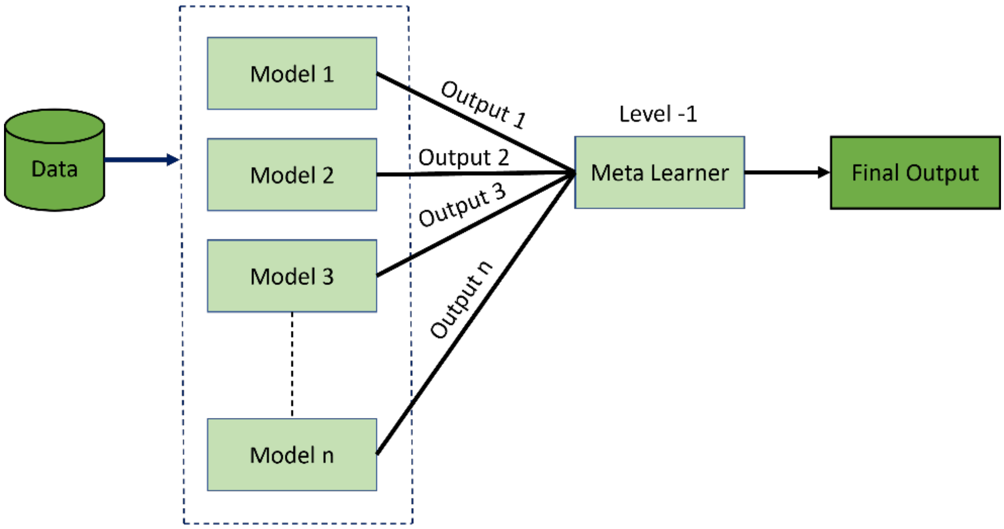


Fig. 2. Meta Ensemble Model

K-Nearest Neighbors (KNN) is one of the algorithms used to classify data based on training data. The data grouping process is based on the number of nearest neighbors (k nearest neighbors). The nearest neighbor search technique that is usually done is to use the Euclidean distance calculation. Euclidean distance is used to find the distance between two points. Equation (1) is a formula for calculating the Euclidean distance [34]. In this study, KNN will be utilized as the meta learner:

$$d = \sqrt{\sum_{i=1}^{m}(x_i - y_i)^2} \tag{1}$$

where, d is a distance, while x and y are features. In this study, we used one nearest k number. Fig. 3 illustrates the process of grouping data based on a predetermined number of K.

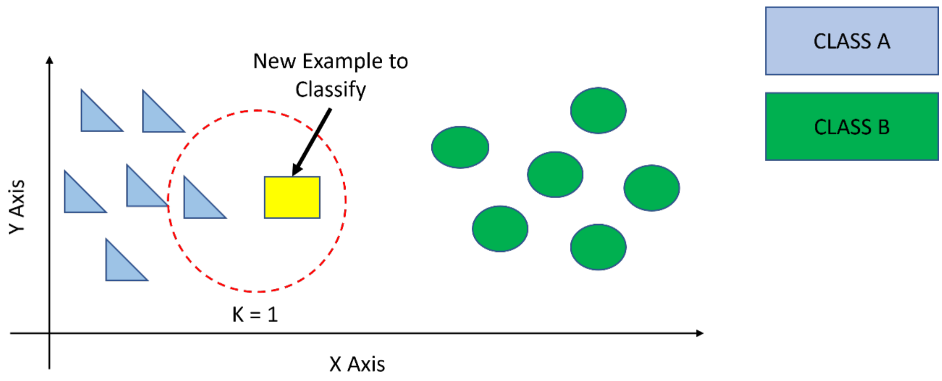


Fig. 3. Illustration of KNN (K-Nearest Neighbors)

2.5. Model Deployment

This study proposes a classification model with a meta ensemble approach using five classification algorithms: Xception, Resnet 15V2, InceptionV3, VGG16, and VGG19 as shown in Fig. 4.

The five algorithms are used after the pre-processing process to classify chest X-ray images. Where each algorithm will produce output in the form of probabilities (Softmax). This study uses the classification results in the form of Softmax in each classification model so that the results are varied. So that the knowledge of the model in classifying data is more diverse. Then the learning process is carried out with the ensemble technique that applies the meta ensemble approach with the KNN classification algorithm. The ensemble technique used in this study also combines the classification results in Softmax. Finally, the KNN algorithm will perform the classification process by selecting the output value that has the closest probability.

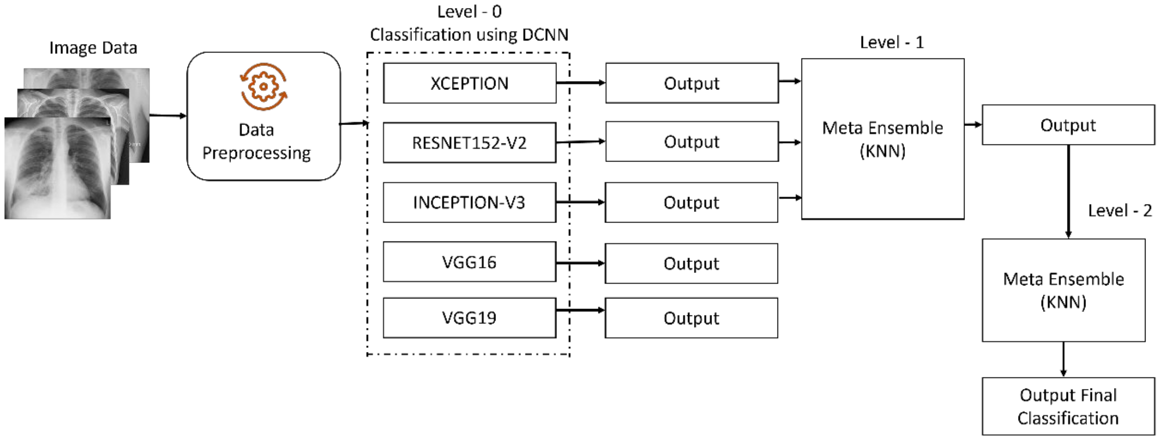


Fig. 4. Schematic Diagram of Model Development

The ensemble stage in this study was carried out using two different concepts, where the first level ensemble was carried out by combining the output of the Xception, ResNet15V2, and InceptionV3 algorithms. Then the output from the first ensemble level is reused for the following learning process, combined with the output from other algorithms, namely VGG16 and VGG19. This process is called ensemble level two. The classification algorithm used at this stage is the same as the previous stage, using KNN as a classification model.

The model learning process is expected to run better using the ensemble model proposed in this study. Such as making the model classification results in the form of probability (Softmax) and using the meta ensemble technique in two levels with the technique of combining the results of the model classification proposed in this study. This is because the classification results produced are in the form of probability (Softmax), and the learning process is carried out more than once using the same classification algorithm, namely KNN in the ensemble process, but using more diverse classification results so that the final classification model is richer in knowledge from the process previous learning.

2.6. Evaluation Matrix

The evaluation matrix used in this study includes accuracy, precision, recall, and F1 scores. The model's classification was evaluated using four metrics: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). A chest X-ray of a pneumonia patient correctly identified as pneumonia by the model is considered a True Positive. If the model correctly identifies a normal chest X-ray (not pneumonia), it is a True Negative. False Positive occurs when the model incorrectly identifies a normal chest X-ray as pneumonia, while False Negative occurs when a chest X-ray of a pneumonia patient is incorrectly identified as normal by the model. Calculations of accuracy, precision, recall, and F1-score values are found in (2), (3), (4), and (5).

$$\text{accuracy} = \frac{\text{TP+TN}}{\text{TN+FP+TP+FN}} \quad (2)$$

$$\text{precision} = \frac{\text{TP}}{\text{FP+ TP}} \quad (3)$$

$$\text{recall} = \frac{\text{TP}}{\text{FN + FP}} \quad (4)$$

$$\text{F1} = 2 \times \frac{\text{recall} \times \text{precision}}{\text{recall+ precision}} \quad (5)$$

3. Results and Discussion

This section describes the experimental stages and analyzes the results of the experiments carried out in this study. The experiment in this study was carried out using the Google Colab Pro platform with GPU resources that allocated 25GB of memory for machine learning projects. Google Colab is an interactive computing document that allows users to write, run, and save programs. It is comparable to a Jupyter notebook and operates on a cloud-based platform accessed through a browser. This study also uses the python programming language with a Keras library to run deep learning algorithm commands and TensorFlow as a library that is used to develop and implement Machine Learning and other algorithms that have many mathematical operations. Based on the experiments that have been carried out, Table 1 contains the results of model accuracy using several models proposed in this study.

Table 1. The Pre-Trained Results for Deep CNN Architectural Model

Pre-Trained Model Architecture	Accuracy	Precision	Recall	F1-Score
Xception	0.97696	0.98	0.99	0.98
ResNet152V2	0.97312	0.99	0.97	0.98
InceptionV3	0.96737	0.98	0.98	0.98
VGG19	0.95393	0.96	0.99	0.98
VGG16	0.96353	0.99	0.95	0.97

The first experiment was initiated by classifying using pre-trained Deep CNN models such as Xception, ResNet153V2, InceptionV3, VGG19, and VGG16. The architecture of the model used in this research is modified first in the layered architecture, which is directly connected from the extractor to the fully connected layer. The optimal result of this research is to use Xception. Because the model proposed in this study applies the ensemble technique, it is necessary to use several algorithms to carry out the ensemble process. Based on the experimental results in Table 1, the classification algorithm chosen for the level 1 ensemble process is the algorithm that has the best classification results. It is hoped that the ensemble model will learn more from models with high accuracy. A learning process will affect the classification results when combined again at level 2 with other models with low accuracy. The next experiment was carried out by comparing the usual meta ensemble method (non-Softmax output), ordinary meta ensemble with Softmax output, multilevel meta ensemble with standard output (non-Softmax output), and multilevel meta ensemble with Softmax output. Table 2 contains the classification results from the experiments carried out.

Table 2. Accuracy Comparison of Original Ensemble Stacking and Multilevel Stacking Ensemble

Architecture	Accuracy	Precision	Recall	F1-Score
Original Ensemble Stacking (using Softmax)	0.9750	0.9800	0.9900	0.9800
Multilevel Ensemble Stacking (using Softmax)	0.9827	0.9800	0.9900	0.9900
Original Ensemble Stacking (non Softmax)	0.9808	0.9800	0.9900	0.9900
Multilevel Ensemble Stacking (non Softmax)	0.9750	0.9800	0.9900	0.9800

This study uses a pre-trained model or a model trained on massive data with weights from ImageNet training so that the pre-trained model has better accuracy. This pre-training model is also included in the Keras library. The model proposed in this study has better accuracy from the trials than the other models, with a test accuracy value of 98.272%. Several things influence this, including the proposed multilevel meta-ensemble technique. This makes the model learn more due to merging the classification model and the relearning technique in stage two. So that makes the model richer in knowledge and makes it easier for the model to find patterns in the classification process. In addition, the classification results using probability (Softmax) also help the model in determining classification because the number of classification classes is more diverse when using probability, so the model studies more variations of classification classes.

Even so, the proposed model has classification results similar to the original stacking technique, whether using Softmax or not, so the ensemble technique using several Deep CNN models can provide good classification results. So that the model performance will be more optimal if it applies the Multilevel Ensemble Stacking technique using Softmax as proposed in this study. In order to test the performance of the proposed classification model, this study made comparisons with other studies that used the similar datasets. Table 3 contains the classification results from the experiments conducted and comparisons with the classification results in other studies.

Table 3. Accuracy comparison of the proposed model in this study with other studies

References	Architecture	Accuracy	Precision	Recall	F1-Score
Muhaza, *et al.* [35]	CNN Model with Oversampling	-	0.9500	0.9600	-
Shagun, *et al.* [36]	Proposed NN with VGG16	0.9215	0.9428	0.9308	0.9540
Enes, *et al.* [37]	Proposed Ensemble CNN	0.9583	-	-	-
Alhassan, *et al.* [38]	Ensemble Model (DenseNet 169, MobileNetV2, and Vision Transformer)	0.9391	0.9396	0.9299	0.9343
Proposed Method	**Multi-Level Meta Stacking using Softmax**	**0.9808**	**0.9800**	**0.9900**	**0.9900**

Based on testing other studies using similar data, this study's proposed multilevel ensemble model has better precision compared to research conducted by Muhaza *et al.* [35] and accuracy, which far beyond the research conducted by Shagun *et al.* [36]. The author also compares with a similar concept, namely ensemble learning, such as research by Enes *et al.* [37] that proposes the ResNet-50, MobileNet, and Xception ensemble models with CNN. However, the model proposed in this study obtains better accuracy. In addition, ensembles using DenseNet 169, MobileNetV2, and Vision Transformer have also been carried out by Alhassan *et al.* [38]. The model proposed in this study obtained better accuracy and

F1 values from the results of these experiments. The confusion matrix is used to evaluate the efficacy of the classification model, as shown in Table 4. In this investigation, the confusion matrix with the highest True Positive Rate (TPR) was obtained, as 383 out of 386 cases were correctly classified for a percentage of 99.222%. The True Negative Rate (TNR) was determined to be 129 out of 135 cases, or 95.555% accuracy, while the False Positive Rate (FPR) was 6 out of 135 cases, or 4.440% error rate. In contrast, the False Negative Rate (FNR) yielded a value of 3 out of 386 cases, or a percentage of 0.777%.

Table 4. Best result of confusion matrix

		Predicted		
		Non-Pneumonia	*Pneumonia*	*Total*
Actual	*Non-Pneumonia*	129 (TN)	6 (FP)	135
	Pneumonia	3 (FN)	383 (TP)	386
	Total	132	389	521

4. Conclusion

The use of the proposed multilevel meta-ensemble architecture and the use of output with probability (Softmax) in this study can affect the model's performance and produce better accuracy. This can help the model explore and learn more words with a more significant number of variations because the ensemble process is carried out in two stages with Softmax. However, the proposed model has the disadvantage that the model training process requires more time. This is due to the use of the deep CNN model architecture and large amounts of data. To further analyze the proposed model architecture in this study, future research can conduct experiments using other classification algorithms, such as machine learning algorithms, or even classification algorithms that do not yet have an excellent ability to recognize image patterns in data. This is intended so that a more in-depth analysis can be carried out regarding the proposed model in this study and whether it can be applied to various classification algorithms to obtain better results.

Declarations

Author contribution. All authors contributed equally to the main contributor to this paper. All authors read and approved the final paper.
Funding statement. None of the authors have received any funding or grants from any institution or funding body for the research.
Conflict of interest. The authors declare no conflict of interest.
Additional information. No additional information is available for this paper.

References

[1] "Pneumonia in children," *World Health Organization*, 2022. Accessed Dec. 01, 2022. [Online]. Available at: https://www.who.int/news-room/fact-sheets/detail/pneumonia.

[2] Kementerian Kesehatan Republik Indonesia," Pneumonia Pada Anak bisa Dicegah dan Diobati", 2020. Accessed May 01, 2022. [Online]. Available at: https://www.kemkes.go.id/.

[3] W. Khan, N. Zaki, and L. Ali, "Intelligent Pneumonia Identification From Chest X-Rays: A Systematic Literature Review," *IEEE Access*, vol. 9, pp. 51747–51771, Jul. 2021, doi: 10.1109/ACCESS.2021.3069937.

[4] O. Stephen, M. Sain, U. J. Maduh, and D.-U. Jeong, "An Efficient Deep Learning Approach to Pneumonia Classification in Healthcare," *J. Healthc. Eng.*, vol. 2019, pp. 1–7, Mar. 2019, doi: 10.1155/2019/4180949.

[5] T. B. Chandra and K. Verma, "Pneumonia Detection on Chest X-Ray Using Machine Learning Paradigm," in *Advances in Intelligent Systems and Computing*, vol. 1022 AISC, Springer Science and Business Media Deutschland GmbH, 2020, pp. 21–33, doi: 10.1007/978-981-32-9088-4_3.

[6] N. Sharma, V. Jain, and A. Mishra, "An Analysis Of Convolutional Neural Networks For Image Classification," *Procedia Comput. Sci.*, vol. 132, pp. 377–384, Jan. 2018, doi: 10.1016/j.procs.2018.05.198.

[7] P. Chagas *et al.*, “Evaluation of Convolutional Neural Network Architectures for Chart Image Classification,” in *2018 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2018, vol. 2018-July, pp. 1–8, doi: 10.1109/IJCNN.2018.8489315.

[8] Z. Wang *et al.*, “Breast Cancer Detection Using Extreme Learning Machine Based on Feature Fusion With CNN Deep Features,” *IEEE Access*, vol. 7, pp. 105146–105158, 2019, doi: 10.1109/ACCESS.2019.2892795.

[9] W. Alakwaa, M. Nassef, and A. Badr, “Lung Cancer Detection and Classification with 3D Convolutional Neural Network (3D-CNN),” *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 8, pp. 66–73, 2017, doi: 10.14569/IJACSA.2017.080853.

[10] J. C. Hung, K.-C. Lin, and N.-X. Lai, “Recognizing learning emotion based on convolutional neural networks and transfer learning,” *Appl. Soft Comput.*, vol. 84, p. 105724, Nov. 2019, doi: 10.1016/j.asoc.2019.105724.

[11] F. Zhuang *et al.*, “A Comprehensive Survey on Transfer Learning,” *Proc. IEEE*, vol. 109, no. 1, pp. 43–76, Jan. 2021, doi: 10.1109/JPROC.2020.3004555.

[12] E. Ayan and H. M. Unver, “Diagnosis of Pneumonia from Chest X-Ray Images Using Deep Learning,” in *2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT)*, Apr. 2019, pp. 1–5, doi: 10.1109/EBBT.2019.8741582.

[13] D. Zhang, F. Ren, Y. Li, L. Na, and Y. Ma, “Pneumonia Detection from Chest X-ray Images Based on Convolutional Neural Network,” *Electronics*, vol. 10, no. 13, p. 1512, Jun. 2021, doi: 10.3390/electronics10131512.

[14] P. Rajpurkar *et al.*, “CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning,” *arXiv Comput. Vis. Pattern Recognit.*, pp. 1–7, Nov. 2017, doi: 10.48550/arXiv.1711.05225.

[15] T. Rahman *et al.*, “Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection Using Chest X-ray,” *Appl. Sci.*, vol. 10, no. 9, p. 3233, May 2020, doi: 10.3390/app10093233.

[16] D. Varshni, K. Thakral, L. Agarwal, R. Nijhawan, and A. Mittal, “Pneumonia Detection Using CNN based Feature Extraction,” in *2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, Feb. 2019, pp. 1–7, doi: 10.1109/ICECCT.2019.8869364.

[17] V. Chouhan *et al.*, “A Novel Transfer Learning Based Approach for Pneumonia Detection in Chest X-ray Images,” *Appl. Sci.*, vol. 10, no. 2, p. 559, Jan. 2020, doi: 10.3390/app10020559.

[18] A. Mabrouk, R. P. Díaz Redondo, A. Dahou, M. Abd Elaziz, and M. Kayed, “Pneumonia Detection on Chest X-ray Images Using Ensemble of Deep Convolutional Neural Networks,” *Appl. Sci.*, vol. 12, no. 13, p. 6448, Jun. 2022, doi: 10.3390/app12136448.

[19] J. R. Ferreira, D. Armando Cardona Cardenas, R. A. Moreno, M. de Fatima de Sa Rebelo, J. E. Krieger, and M. Antonio Gutierrez, “Multi-View Ensemble Convolutional Neural Network to Improve Classification of Pneumonia in Low Contrast Chest X-Ray Images,” in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. 2020, vol. 2020-July, pp. 1238–1241, doi: 10.1109/EMBC44109.2020.9176517.

[20] “Chest X-Ray Images (Pneumonia),” *Kaggle*. Accessed May 02, 2018. [Online]. Available at: https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia.

[21] D. S. Kermany *et al.*, “Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning,” *Cell*, vol. 172, no. 5, pp. 1122-1131.e9, Feb. 2018, doi: 10.1016/j.cell.2018.02.010.

[22] H. Panwar, P. K. Gupta, M. K. Siddiqui, R. Morales-Menendez, and V. Singh, “Application of deep learning for fast detection of COVID-19 in X-Rays using nCOVnet,” *Chaos, Solitons & Fractals*, vol. 138, p. 109944, Sep. 2020, doi: 10.1016/j.chaos.2020.109944.

[23] H. Polat and H. Danaei Mehr, “Classification of Pulmonary CT Images by Using Hybrid 3D-Deep Convolutional Neural Network Architecture,” *Appl. Sci.*, vol. 9, no. 5, p. 940, Mar. 2019, doi: 10.3390/app9050940.

[24] A. Mortazi and U. Bagci, “Automatically Designing CNN Architectures for Medical Image Segmentation,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture*

Notes in Bioinformatics), vol. 11046 LNCS, Springer Verlag, 2018, pp. 98–106, doi: 10.1007/978-3-030-00919-9_12.

[25] C. Szegedy *et al.*, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2015, vol. 07-12-June, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.

[26] M. I. Sarker, H. Kim, D. Tarasov, and D. Akhmetzanov, "Inception Architecture and Residual Connections in Classification of Breast Cancer Histology Images," *arXiv Image Video Process.*, vol. 1–8, Dec. 2019, doi: 10.48550/arXiv.1912.04619.

[27] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, vol. 2017-Janua, pp. 1800–1807, doi: 10.1109/CVPR.2017.195.

[28] S. H. Kassani, P. H. Kassani, R. Khazaeinezhad, M. J. Wesolowski, K. A. Schneider, and R. Deters, "Diabetic Retinopathy Classification Using a Modified Xception Architecture," in *2019 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, Dec. 2019, pp. 1–6, doi: 10.1109/ISSPIT47144.2019.9001846.

[29] A. Sengupta, Y. Ye, R. Wang, C. Liu, and K. Roy, "Going Deeper in Spiking Neural Networks: VGG and Residual Architectures," *Front. Neurosci.*, vol. 13, p. 95, Mar. 2019, doi: 10.3389/fnins.2019.00095.

[30] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, Sep. 2014, pp. 1–14, doi: 10.48550/arXiv.1409.1556.

[31] D. Sarwinda, R. H. Paradisa, A. Bustamam, and P. Anggia, "Deep Learning in Image Classification using Residual Network (ResNet) Variants for Detection of Colorectal Cancer," *Procedia Comput. Sci.*, vol. 179, pp. 423–431, Jan. 2021, doi: 10.1016/j.procs.2021.01.025.

[32] Z.-H. Zhou, "Ensemble Learning," in *Machine Learning*, Singapore: Springer Singapore, 2021, pp. 181–210, doi: 10.1007/978-981-15-1967-3_8.

[33] A. Almasri, E. Celebi, and R. S. Alkhawaldeh, "EMT: Ensemble Meta-Based Tree Model for Predicting Student Performance," *Sci. Program.*, vol. 2019, pp. 1–13, Feb. 2019, doi: 10.1155/2019/3610248.

[34] W. Xing and Y. Bei, "Medical Health Big Data Classification Based on KNN Classification Algorithm," *IEEE Access*, vol. 8, pp. 28808–28819, 2020, doi: 10.1109/ACCESS.2019.2955754.

[35] R. Yacouby and D. Axman, "Probabilistic Extension of Precision, Recall, and F1 Score for More Thorough Evaluation of Classification Models," in *Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems*, Nov. 2020, pp. 79–91, doi: 10.18653/v1/2020.eval4nlp-1.9.

[36] M. Liebenlito, Y. Irene, and A. Hamid, "Classification of Tuberculosis and Pneumonia in Human Lung Based on Chest X-Ray Image using Convolutional Neural Network," *Inpr. Indones. J. Pure Appl. Math.*, vol. 2, no. 1, pp. 24–32, Mar. 2020, doi: 10.15408/inprime.v2i1.14545.

[37] S. Sharma and K. Guleria, "A Deep Learning based model for the Detection of Pneumonia from Chest X-Ray Images using VGG-16 and Neural Networks," *Procedia Comput. Sci.*, vol. 218, pp. 357–366, Jan. 2023, doi: 10.1016/j.procs.2023.01.018.

[38] E. Ayan, B. Karabulut, and H. M. Ünver, "Diagnosis of Pediatric Pneumonia with Ensemble of Deep Convolutional Neural Networks in Chest X-Ray Images," *Arab. J. Sci. Eng.*, vol. 47, no. 2, pp. 2123–2139, Feb. 2022, doi: 10.1007/s13369-021-06127-z.