IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 13, No.4, December 2024, pp. 4169~4177
ISSN: 2252-8938, DOI: 10.11591/ijai.v13.i4.pp4169-4177

Pneumonia detection on x-ray image using improved depthwise separable convolutional neural networks

Islam Nur Alam[1], Ghinaa Zain Nabiilah[1], Erna Fransisca Angela Sihotang[2], Bakti Amirul Jabar[1]
[1]Department of Computer Science, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia
[2]Department of Statistics, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia

Article Info

Article history:

Received Nov 13, 2023
Revised Apr 5, 2024
Accepted Apr 17, 2024

Keywords:

Chest x-ray
Convolutional neural network
MobileNetV2
Pneumonia
Xception

ABSTRACT

A single neural network model cannot capture intricate and diverse features due to its ability to learn only a finite set of patterns from the data. Additionally, training and utilising a single model can be computationally demanding. Experts propose incorporating multiple neural network models to address these constraints to extract complementary attributes. Previous research has highlighted challenges network models face, including difficulties in effectively capturing highly detailed spatial features, redundancy in network structure parameters, and restricted generalisation capabilities. This study introduces an innovative neural network architecture that combines the Xception module with the inverse residue structure to tackle these issues. Considering this, the paper presents a model for detecting pneumonia in X-ray images employing an improved depthwise separable convolutional network. This network architecture integrates the inverse residual structure from the MobileNetV2 model, using the rectified linear unit (ReLU) non-linear activation function throughout the entire network. The experimental results show an impressive recognition rate with a test accuracy of 97.24% on the chest x-ray dataset. This method can extract more profound and abstract image features while mitigating overfitting issues and enhancing the network's generalisation capacity.

Corresponding Author:

Islam Nur Alam
Department of Computer Science, School of Computer Science, Bina Nusantara University
Jakarta 11480, Indonesia
Email: islam.alam@binus.ac.id

1. INTRODUCTION

Pneumonia, an infection of the lungs causing inflammation and fluid buildup, poses a significant global health threat. This condition, characterized by shortness of breath and reduced oxygen intake, claims the lives of countless children under five annually. In 2019 alone, worldwide figures show the death toll reaching 740,180, with 314,455 of those young victims residing in Indonesia [1], [2]. These images can be ambiguous, leading to misdiagnosis or conflicting interpretations, even for experienced radiologists. Furthermore, manual analysis is time-consuming, potentially delaying critical treatment [3].

To address the challenges of interpreting chest X-rays, the implementation of a computerized system emerges as a crucial solution to assist radiologists in identifying pneumonia. Machine learning, particularly adept at image classification, exhibits promising potential in this domain. Chandra and Verma [4] exemplifies this, where they evaluated the efficacy of various machine learning classifiers, including logistic regression, multilayer perceptron, random forest, and sequential minimal optimization, in detecting pneumonia from X-ray images [4]. Their findings provide compelling evidence of the capabilities of these models, laying the groundwork for further exploration and optimization within this promising field.

Beyond aiding diagnosis, the study employs feature extraction prior to classification, ultimately achieving an impressive 95.53% accuracy with logistic regression, highlighting the potential of machine learning approaches. Nevertheless, limitations arise when handling significant data volumes, as evidenced by the study's use of a relatively small dataset (412 images) [5]. Deep learning offers a promising alternative for overcoming these limitations, particularly in image classification domains like medical imaging. Convolutional neural networks (CNNs) stand out as a popular and effective deep learning method [6], [7]. Their success in detecting various medical issues like breast and lung cancer, brain tumors, and skin diseases underscores their vast potential for pneumonia detection as well [8].

Leveraging transfer learning, a technique utilizing knowledge gained from one problem to solve another, holds immense potential for addressing pneumonia detection challenges [9]. This approach capitalizes on similarities between problems, accelerating learning in scenarios where acquiring data is difficult or costly, like in medical contexts [10]. At its core, transfer learning bridges the gap between familiar and unfamiliar information, fostering new insights. It entails transferring knowledge from a well-known "source domain" (e.g., general image classification) to the unfamiliar "target domain" (pneumonia detection in X-rays). The aim is to explore effective methods for this knowledge transfer, allowing established models to apply previous learnings to new information effectively. Within machine learning, transfer learning categorizes into four approaches based on their methods: feature-based, model-based, relationship-based, and sample-based [11].

In the realm of deep neural networks, the training of models with a multilayer architecture using a substantial amount of data is imperative for acquiring practical features and enhancing recognition accuracy. While earlier studies have explored various aspects of deep learning architectures, critical examination reveals certain gaps that need addressing. Specifically, the inadequacy of single network models in extracting intricate and comprehensive features has been acknowledged. Additionally, the expansive structure of networks, characterized by numerous parameters, poses challenges related to computational resources and efficiency. This paper aims to address these gaps by introducing an enhanced Xception network model specifically tailored for pneumonia detection in X-ray images.

The Xception model, conceived as an enhancement to Google's Inception-v3 architecture, employs depthwise and pointwise convolutions to efficiently extract information from diverse channels and convolution kernels. This innovative approach significantly reduces the number of parameters and associated computational expenses. Furthermore, the modified Xception network structure incorporates the inverted residual design from MobileNetV2, effectively mitigating issues related to gradient disappearance and explosion while enhancing gradient propagation between layers. Recognizing the limitations inherent in neural network model training, such as limited data availability and the risk of overfitting, our study employs data augmentation techniques to augment the image dataset. However, it is important to acknowledge that the model outlined in this study is not without its limitations, including suboptimal recognition rates for images afflicted by significant noise, low resolution, and severe occlusion. Future research endeavors will continue to address these concerns, contributing to the ongoing refinement and advancement of pneumonia detection using deep neural networks.

2. RELATED WORK

Building upon successful implementations of deep learning for pneumonia detection in chest X-ray images, several studies have demonstrated promising advancements. Ayan and Ünver [12] achieved notable results by leveraging transfer learning, fine-tuning, and data augmentation with adapted Xception and VGG16 models. Their work showcased the superior performance of the Xception model compared to VGG16. In 2021, Zhang *et al*. [13] modified the VGG architecture, demonstrating its potential competitiveness among established models like VGG-16, RES-50, Xception, DenseNet21, and MobileNet. These efforts highlight the ongoing refinement of deep learning approaches for accurate pneumonia detection.

While prior studies achieved impressive results, our current study aims to explore ensemble stacking, a technique that combines predictions from multiple models. Leveraging powerful architectures like Xception, Resnet152V2, InceptionV3, VGG16, and VGG19, we have introduced a modification named multilevel ensemble stacking. The primary goal is to further enhance the accuracy of diagnosing pneumonia through chest X-rays.

Recent years have witnessed various methods for identifying pneumonia from chest X-ray images, primarily using deep learning or deep CNN approaches. Rajpurkar *et al.* [14] employed a 121-layer CNN model, achieving success in detecting pneumonia among 14 different chest-related diseases. Rahman *et al*. [15] utilized transfer learning with four deep learning algorithms, where DenseNet201 outperformed others, achieving a 98% accuracy rate. Varshni *et al*. [16] explored alternative machine learning methodologies like support vector machine (SVM), naive Bayes, k-nearest neighbors (KNN), and random forest in conjunction with deep CNN models. The study highlighted DenseNet-169 with SVM, yielding an area under the ROC curve (AUC) of 0.8002 [16].

In the realm of image classification, the latest research endeavors focus on enhancing accuracy through ensemble learning, combining top-performing models. Chouhan *et al*. [17] achieved an accuracy rate of 96.4% by combining deep CNN models, while Mabrouk *et al*. [18] merged vision transformer, MobileNetV2, and DenseNet169 for an optimal accuracy of 93.91%. The model we designed introduces improvements by leveraging concepts from GoogLeNet, Xception, and ResNet. We combine the Xception model with the inverted residual structure, recognized as an efficient and influential deep learning framework. Additionally, the use of the global average pooling layer at the end of the model contributes to enhancing its accuracy.

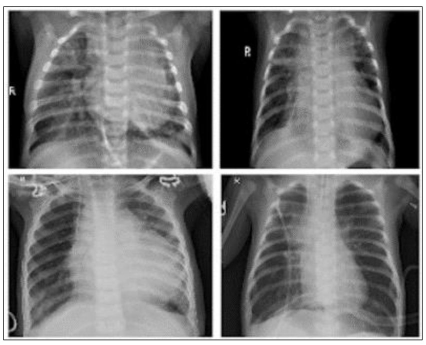
3. METHOD

The study begins by collecting data in the form of "jpg" formatted X-ray images. The next step involves data preprocessing using image augmentation techniques, which are detailed in the "dataset setup" chapter. Next, we construct a deep neural network model known as depthwise separable CNNs. The strategy we apply to this model aims to separate channel and spatial correlations, with the goal of saving network parameters and improving model performance. Initially, the network employs a depthwise-pointwise convolution structure, where depthwise convolution is performed first, followed by pointwise convolution in the second step. Subsequently, we conduct experiments by performing several initial hyperparameter adjustments, namely batch size of 32, 100 epochs, and using the Adam optimizer with a learning rate of 0.001. Once the experiments are completed, we evaluate the model and compare it with previous research through a benchmarking process.

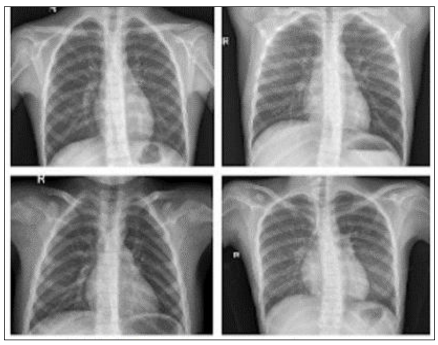
3.1. Datasets setup

The present investigation utilised a publicly available dataset developed by Mooney and obtained via the online platform Kaggle [18]. The dataset comprises roughly 5,197 chest X-ray images, which have been classified into two distinct categories: pneumonia and normal. The chest X-ray pictures utilised in this investigation were selected from a cohort of prior instances involving paediatric patients aged one to five who had received chest X-ray examinations as part of their routine clinical evaluations at the Guangzhou women and children's medical centre in Guangzhou. The images were acquired via the anterior-posterior view, the conventional method employed for obtaining images of this nature [19].

To guarantee the reliability of the studied data, a meticulous quality control screening process was applied to the chest X-ray images, filtering out low-quality or incomprehensible ones. Two expert physicians then independently evaluated the remaining images to validate the accuracy of the AI system's training data. Additionally, a third expert assessed the dataset to minimise potential grading errors. In order to enhance the dataset, a data augmentation technique was also employed for all of the datasets [20]. The augmented chest X-ray dataset includes 5,948 train set images and 1,487 test set images. This collection features chest X-ray images from patients with pneumonia and those without the condition. An example image from the dataset is depicted in Figures 1(a) and 1(b).



(a)



(b)

Figure 1. An illustration of a dataset image is provided as sample data of (a) pneumonia; total data: 4273 images and (b) non pneumonia; total data: 3162 images

3.2. Construction of the neural network model

Equations to streamline the network's weight count, simplify the model's structure, and enhance accuracy, this paper's network model lessens parameters through partial connectivity and weight sharing, displaying hierarchical expression characteristics. It integrates the conceptual structure of depthwise separable convolution. In the Xception network module, the process begins with a 1×1 convolution applied to the input image. Subsequently, a 3×3 convolution is employed on each channel after the convolution stage, leading to the consolidation of outcomes. In contrast to the Inception-v3 [21] network model, this approach amplifies model efficiency without inflating complexity. Moreover, a constructed network model applies a residual connection mechanism to address performance degradation and gradient vanishing issues, enabling the training of deeper networks while upholding optimal performance. The schematic representation of the architectural arrangement of the neural network with reduced weight is illustrated in Figure 2. Image data enters from the left, progresses through the middle layer to extract features, and culminates in classification results through the softmax function.

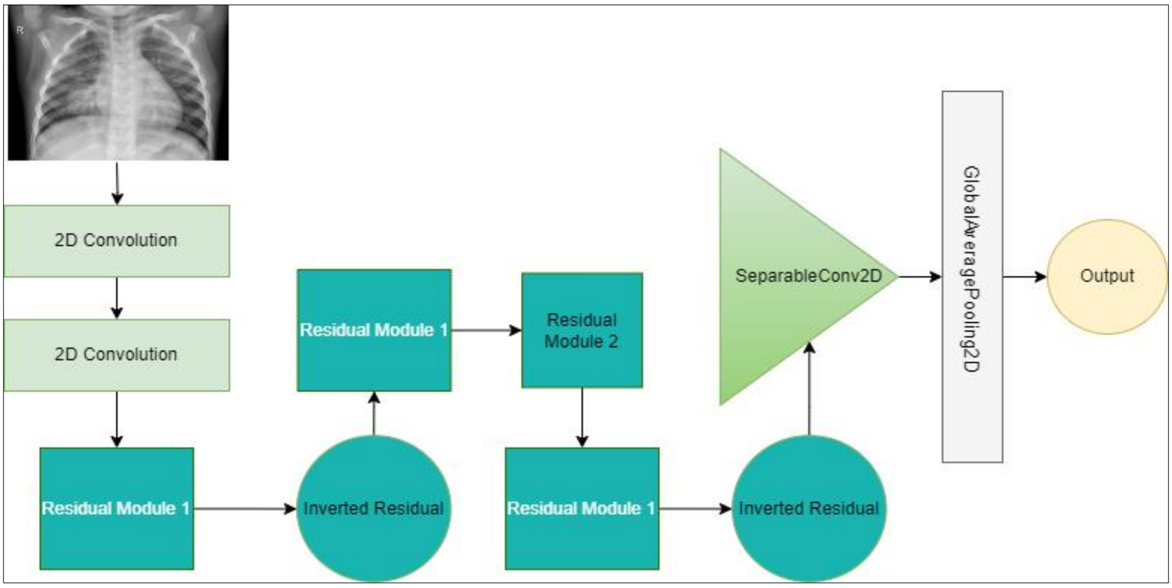


Figure 2. Neural network architecture diagram

Convolution is a way to extract features from images using a CNN. It works by sliding a small filter, called a kernel, over the image and multiplying each pixel value in the kernel with its corresponding pixel value. The sum of the products is then assigned to a new image called the feature map. Convolution can be used to enhance the critical features of an image and reduce noise. It can also be used to reduce an image's size, making it easier for the CNN to process. There are two types of convolution padding: valid and same. Valid padding does not add any padding to the image, so the feature map is smaller than the original image [22]. The same padding adds padding to the image so that the feature map is the same size as the original image. The convolution operation is computed by the (1), where X_j^L is the j -th feature map unit of the L -th layer, X_i^{L-1} is the i -th input of the $L-1$ -th layer, θ_{ij} Represents the convolution kernel, b is the bias unit, and $g(x)$ is the activation function.

$$X_j^L = g\left(\sum_i X_i^{L-1} \theta_{ij} + b\right) \tag{1}$$

The network model uses max pooling after the depthwise separable convolution layer to compress features and extract the most important ones, simplifying the network structure and reducing the risk of overfitting. Max pooling works by selecting the maximum value from a small region of the input feature map [23]. This work employs global average pooling as a substitution for the fully connected layer to mitigate overfitting and decrease the parameter count. Global average pooling involves aggregating the input's spatial information, enhancing the network's resilience to spatial alterations [24], [25].

3.3. Depthwise separable convolution

The construction follows a strategy separating channel and spatial correlations to save network parameters and improve network model performance. Initially, the network employs a depthwise-pointwise convolution structure, undertaking depthwise convolution first and then another in the second step. These dual convolution operations serve distinct functions, extracting features from diverse channels with varying convolution kernels. Maintaining the data integrity, the two convolution actions have no interceding nonlinearity (ReLu), as presented in Figure 3. Initially, a 1×1 channel correlation convolution is executed, followed by a 3×3 convolution with the same output channels. This two-step feature extraction technique minimises the space and time expenses in constructing and training the network while achieving image feature extraction comparable to standard convolution. It significantly reduces computational load while upholding neural network accuracy.

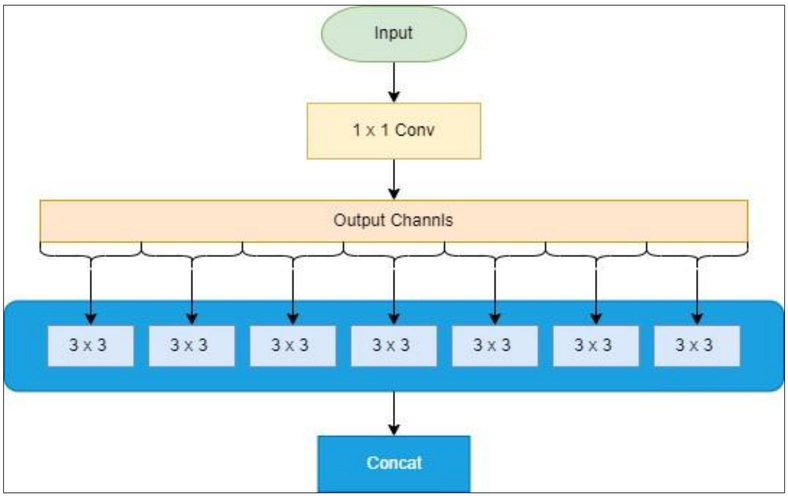


Figure 3. Decoupling cross-channel and spatial correlation

3.4. Network architecture

This paper presents an improved technique for detecting pneumonia in X-ray images through a depthwise separable convolutional network. The network's design integrates feature recognition inspired by the structure of Xception [26]. They mainly emphasise the consecutive handling of features from three channels after their fusion to improve the precision of detecting pneumonia. The network's structure and detailed parameter information are depicted in Figure 2 and Table 1. Initially, it entails successive 2D convolution procedures spanning two layers. The convolution kernel is 3×3, featuring 32 and 64 individual convolution kernels with a stride of 1. Following this, the output experiences a series of stages involving residual module 1, an inverted residual microstructure, residual module 1, residual module 2, and a subsequent sequence of residual module 1, an inverted residual microstructure. The quantity of convolution kernels in these structures varies: 128, 256, 256, 364, 364, and 512. The depthwise separable convolution employs a 3×3 kernel, while MaxPooling uses a 3×3 kernel, and the 2D convolution layer uses a 1×1 kernel. Residual module 1 operates with a step size of 2, whereas the others use a step of 1. The output moves through a depthwise separable convolution layer with a 3×3 kernel, 728 kernels, and a step size 1. Eventually, the output undergoes a global average pooling layer and Softmax classifier. All 2D convolution and depthwise separable convolution operations pass through batch normalization and ReLU activation layers to expedite the convergence of the network structure and augment the extraction of non-linear features.

Table 1. Network structure and parameter table

Input	Layer	*c*	*s*
224×224×3	Conv2D	32	1
224×224×32	Conv2D	64	1
112×112×64	Residual module 1	128	2
28×28×128	Inverted residual structure × 4	256	1
28×28×256	Residual module 1	256	2
14×14×256	Residual module 2	364	1
14×14×364	Residual module 1	364	2
7×7×364	Inverted residual structure × 2	512	1
7×7×512	SeparableConv2D	728	1
7×7×728	GlobalAveragePooling2D	-	-
1×1×728	Conv2D	2	-

4. RESULTS AND DISCUSSION

This research centres on detecting pneumonia from static X-ray images. The investigation employed a batch size of 32, an initial learning rate set at 0.001, and utilised the Adam optimiser to refine the training process. Additionally, the network incorporated a width multiplier, α, set to 1. This integration of parameter α aims to reduce the number of network parameters, a strategy observed to yield more favourable results than reducing the model's depth.

Based on the findings in Figure 4, it can be seen that the network model is constructed by combining the Xception model and the inverse residual structure correlated with separable CNN. The method proposed in this study tends to have an accuracy proportion reaching 0.9724, which is much higher than the method used by ensemble model: DenseNet 169, MobileNetV2, and vision transformer. Our study demonstrates that network models with stronger robustness and higher generalization ability are not necessarily correlated with poor performance on deep neural networks. The proposed method can leverage the improved Xception network model without negatively impacting the single network model. Although models with more complex and rich features, and very large network structures, have many parameters and consume more space resources, this method is able to reduce the number of parameters and computational costs.

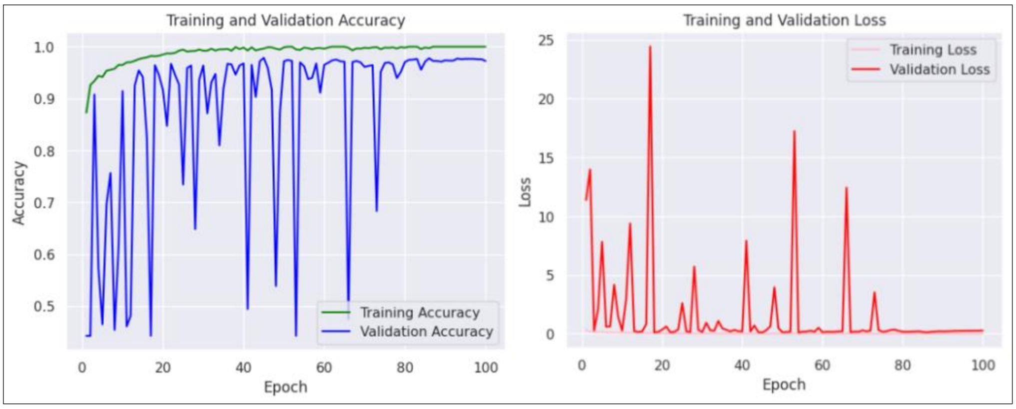


Figure 4. The training and testing accuracy and loss with chest X-ray images dataset

This study comprehensively explores the integration of the Xception model with an inverted residual structure introduced into the network architecture. Integrating this model with the improved Xception network structure provides a solution to the challenges of gradient vanishing and gradient explosion, while also enhancing the gradient propagation capability between product layers. Nevertheless, some shortcomings were found in the model built in this paper, such as low recognition rate for images with large noise, low resolution, and serious occlusion. Therefore, further in-depth studies may be needed to confirm these findings, especially in terms of image processing using specific methods before entering the feature extraction process by the architecture. This is to further save the computation process and sharpen the image analysis.

Furthermore, derived from the findings of the confusion matrix chart depicted in Figure 5, a significant element enhancing the model's performance to achieve precise outcomes comprises the true positive (TP) and true negative (TN) values. TP refers to the volume of positive data accurately identified by the model. TN represents the volume of harmful data correctly classified as unfavourable by the model. The more TP or TN the model has, the higher the accuracy of the model. Both TP and TN values positively contribute to accuracy, as they represent correct predictions by the model.

4.1. Comparative experiments of different methods

Table 2 provides a comparative analysis of accuracy for various models developed by different researchers. Among the models mentioned, the paper's proposed model stands out with the highest accuracy score of 97.24%. This implies that the model outperforms the alternatives presented in the table, including a CNN model with oversampling, a proposed neural network with VGG16, a proposed ensemble CNN, and an ensemble model using various architectures like DenseNet 169, MobileNetV2, and vision transformer. The notable difference in accuracy signifies that the paper's model has substantially advanced in the field of study, likely achieving better results. It attests to the effectiveness of the "improved depthwise separable convolutional network" in achieving higher accuracy, a crucial metric in various machine learning and computer vision applications.

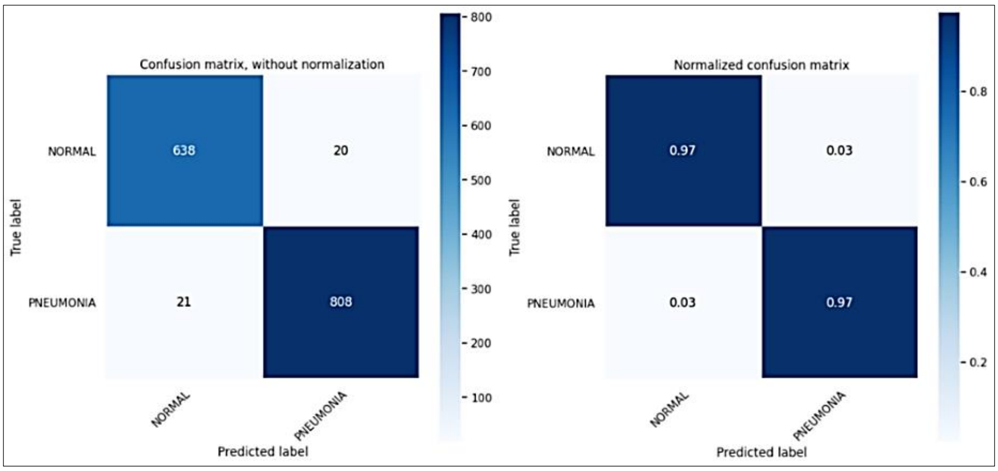


Figure 5. Confusion matrix on chest x-ray images dataset

Table 2. Comparison of test accuracy with previous research

Reference	Method	Accuracy (%)
Liebenlito *et al*. [27]	CNN Model with oversampling	95.00
Sharma and Guleria [28]	Proposed NN with VGG16	92.15
Ayan *et al*. [29]	Proposed ensemble CNN	95.83
Mabrouk *et al*. [18]	Ensemble Model: DenseNet 169, MobileNetV2, and vision transformer	93.91
Paper Method	Improved depthwise separable convolutional network	97.24

5. CONCLUSION

This research leverages the depthwise separable convolutional network architecture, which integrates a residual network with an inverse residual structure. By addressing limitations inherent in conventional algorithms for pneumonia detection in X-ray images, such as the inability to extract high-level depth features and the weak generalization ability of a single deep network model, this paper introduces an innovative method for pneumonia detection based on depth-resolved enhanced convolutional networks. The proposed approach adeptly combines traditional feature extraction methodologies with artificial neural networks, enabling the capture of more profound and abstract image features. This results in a reduction in the impact of lighting and variations in lung images, among other factors. Through the integration of the Xception model with the inverse residual structure, the network model successfully mitigates overfitting, addressing issues related to gradient loss and overamplification. Empirical findings demonstrate the specific enhancement of pneumonia detection accuracy by the proposed model, reinforcing the network's resilience and generalization ability. Suggestions for future research include exploring the model's susceptibility to more complex variations and occlusions, as well as more realistic clinical scenarios, such as variations in patient age and health conditions, and the possibility of X-ray images having different levels of damage. Furthermore, the study can be extended to consider larger datasets, incorporating data from diverse sources and origins to enhance the model's generalizability beyond the training dataset. Additionally, to deepen the understanding of the characteristics of X-ray images, considering more detailed analysis techniques, such as image segmentation for infected area recognition, could be the next step in advancing this research.

ACKNOWLEDGEMENTS

The authors express their gratitude to the Department of Computer Science, School of Computer Science Bina Nusantara University for their appreciated encouragement.

REFERENCES

[1] K. Kanwal, S. G. Khalid, M. Asif, F. Zafar, and A. G. Qurashi, "Diagnosis of Community-Acquired pneumonia in children using photoplethysmography and machine learning-based classifier," *Biomed Signal Process Control*, vol. 87, Jan. 2024, doi: 10.1016/j.bspc.2023.105367.

[2] A. M. Sobirovna, "Causes of pneumonia in children," *Science and Innovation: International Scientific Journal*, vol. 3, 2024, doi: 10.5281/zenodo.10598855.

[3] O. Stephen, M. Sain, U. J. Maduh, and D.-U. Jeong, "An efficient deep learning approach to pneumonia classification in healthcare," *Journal of Healthcare Engineering*, vol. 2019, pp. 1–7, Mar. 2019, doi: 10.1155/2019/4180949.

[4] T. B. Chandra and K. Verma, “Pneumonia detection on chest x-ray using machine learning paradigm*,” 3rd International Conference on Computer Vision and Image Processing: CVIP 2018*, 2020, pp. 21–33, doi: 10.1007/978-981-32-9088-4_3.
[5] N. Sharma, V. Jain, and A. Mishra, “An analysis of convolutional neural networks for image classification,” *Procedia Computer Science*, vol. 132, pp. 377–384, 2018, doi: 10.1016/j.procs.2018.05.198.
[6] P. Chagas *et al.*, “Evaluation of convolutional neural network architectures for chart image classification,” in *2018 International Joint Conference on Neural Networks (IJCNN)*, IEEE, Jul. 2018, pp. 1–8, doi: 10.1109/IJCNN.2018.8489315.
[7] 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.
[8] W. Alakwaa, M. Nassef, and A. Badr, “Lung cancer detection and classification with 3D convolutional neural network (3D-CNN),” *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 8, 2017, doi: 10.14569/IJACSA.2017.080853.
[9] I. N. Alam, I. H. Kartowisastro, and P. Wicaksono, “Transfer learning technique with efficientnet for facial expression recognition system,” *Revue d’Intelligence Artificielle*, vol. 36, no. 4, pp. 543–552, Aug. 2022, doi: 10.18280/ria.360405.
[10] J. C. Hung, K.-C. Lin, and N.-X. Lai, “Recognizing learning emotion based on convolutional neural networks and transfer learning,” *Applied Soft Computing*, vol. 84, Nov. 2019, doi: 10.1016/j.asoc.2019.105724.
[11] F. Zhuang *et al.*, “A comprehensive survey on transfer learning,” *Proceedings of the 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)*, IEEE, 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, Jul. 2021, doi: 10.3390/electronics10131512.
[14] P. Rajpurkar *et al.*, “CheXNet: Radiologist-level pneumonia detection on chest x-rays with deep learning,” *arXiv-Computer Science*, 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,” *Applied Sciences*, vol. 10, no. 9, 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)*, IEEE, 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,” *Applied Sciences*, vol. 10, no. 2, p. 559, Jan. 2020, doi: 10.3390/app10020559.
[18] A. Mabrouk, R. P. D. Redondo, A. Dahou, M. A. Elaziz, and M. Kayed, “Pneumonia detection on chest x-ray images using ensemble of deep convolutional neural networks,” *Applied Sciences*, vol. 12, no. 13, Jun. 2022, doi: 10.3390/app12136448.
[19] D. S. Kermany *et al.*, “Identifying medical diagnoses and treatable diseases by image-based deep learning,” *Cell*, vol. 172, no. 5, pp. 1122-1131, Feb. 2018, doi: 10.1016/j.cell.2018.02.010.
[20] C. S. Won, “Multi-scale CNN for fine-grained image recognition,” *IEEE Access*, vol. 8, pp. 116663–116674, 2020, doi: 10.1109/ACCESS.2020.3005150.
[21] M. Lin, Q. Chen, and S. Yan, “Network in network,” *arXiv-Computer Science,* pp. 1-10, Dec. 2013, doi: 10.48550/arXiv.1312.4400.
[22] F. Chollet, “Xception: deep learning with depthwise separable convolutions,” *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 1800-1807, doi: 10.1109/CVPR.2017.195.
[23] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “MobileNetV2: inverted residuals and linear bottlenecks,” *arXiv-Computer Science*, pp. 1-14, Jan. 2018, doi: 10.48550/arXiv.1801.04381.
[24] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *2016 IEEE Conference on Computer Vision and Pattern Recognition Deep*, 2016, doi: 10.1109/CVPR.2016.90.
[25] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, “Inception-v4, inception-resnet and the impact of residual connections on learning,” *Thirty-First AAAI Conference on Artificial Intelligence*, vol. 3, no. 1, 2017, doi: 10.1609/aaai.v31i1.11231.
[26] S. Li, H. Qu, X. Dong, B. Dang, H. Zang, and Y. Gong, “Leveraging deep learning and xception architecture for high-accuracy MRI classification in alzheimer diagnosis,” *arXiv-Electrical Engineering and Systems Science*, pp. 1-9, Mar. 2024, doi: 10.48550/arXiv.2403.16212.
[27] 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,” *InPrime: Indonesian Journal of Pure and Applied Mathematics*, vol. 2, no. 1, pp. 24–32, Mar. 2020, doi: 10.15408/inprime.v2i1.14545.
[28] 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 Computer Science*, vol. 218, pp. 357–366, 2023, doi: 10.1016/j.procs.2023.01.018.
[29] E. Ayan, B. Karabulut, and H. M. Ünver, “Diagnosis of pediatric pneumonia with ensemble of deep convolutional neural networks in chest x-ray images,” *Arabian Journal for Science and Engineering*, vol. 47, no. 2, pp. 2123–2139, Feb. 2022, doi: 10.1007/s13369-021-06127-z.

BIOGRAPHIES OF AUTHORS

Islam Nur Alam is a lecturer at Bina Nusantara University (BINUS). He has two years of experience as a data science researcher with a proven track record in building successful algorithms and predictive models for image classification using convolutional neural networks. He is highly proficient in clustering and classification, content-based filtering, data analysis and visualisation. He also continues to hone individual skills in data science, mainly focusing on natural language processing and machine translation tasks. He can be contacted at email: islam.alam@binus.ac.id.

Ghinaa Zain Nabiilah is a lecturer from Bina Nusantara University (BINUS). She graduated from BINUS, Department of Computer Science in 2023. Since 2020, her research has been related to natural language processing, especially in analysing human personality and emotions. In addition, she is also active in research on the management and investigation of toxic sentences, hoaxes, and hate speech for decision support. She can be contacted at email: ghinaa.nabiilah@binus.ac.id.

Erna Fransisca Angela Sihotang is a lecturer in Department of Statistics at School of Computer Science in Bina Nusantara University. She finished her Bachelor's in Statistics at Diponegoro University and then continued her Master's in Computer Science at BINUS University. Her research interests are applied statistics, decision model, and statistical computing. She can be contacted at email: erna.sihotang@binus.ac.id.

Bakti Amirul Jabar is an accomplished academic and researcher with a Bachelor of Science degree in Mathematics from Institut Teknologi Bandung and a Master's degree in Computer Science from BINUS University. He currently serves as a lecturer at BINUS University, where he imparts his knowledge in artificial intelligence. His research contributions to AI have earned him a prominent place in the academic community, and he is actively involved in advancing the understanding and application of AI. For inquiries or collaboration. He can be contacted at email: bakti.jabar@binus.ac.id.