

Combining depthwise separable convolution and transfer learning for pneumonia classification

Puspita Kartikasari¹, Rukun Santoso¹, Suparti¹, Rizwan Arisandi², Vikri Haikal¹

¹Department of Statistics, Faculty of Science and Mathematics, Universitas Diponegoro, Semarang, Indonesia

²Department of Computer Science, School of Computer Science, Universitas Bina Nusantara, Jakarta, Indonesia

Article Info

Article history:

Received Aug 6, 2025

Revised Aug 3, 2026

Accepted Aug 12, 2026

Keywords:

Depthwise separable convolution

Particle swarm optimization

Pneumonia detection

Pneumonia imaging

Soft voting

Transfer learning

ABSTRACT

Pneumonia represents a prevalent pulmonary illness within the general population, significantly impacting diverse demographic groups, most notably infants. Therefore, identifying the disease at an early stage is essential to manage the condition effectively and boost the likelihood of a complete recovery via proper therapeutic interventions. This research endeavors to construct a robust early diagnostic framework based on medical imaging of pneumonia. The methodology introduced evaluates the synergy of depthwise separable convolution, transfer learning, and soft voting mechanisms, where each predictive model undergoes rigorous optimization via particle swarm optimization (PSO). For this study, the dataset is strictly categorized into two distinct groups: normal and pneumonia-infected. The analytical findings demonstrate that the PSO-optimized soft voting ensemble achieves a peak accuracy rate of 92.11%. Such a substantial accuracy level suggests that the developed framework can function as a dependable supplementary diagnostic instrument in clinical environments, which could minimize human diagnostic errors and substantially reduce the time radiologists spend screening critical cases. Moreover, embedding these computational methodologies into interactive, user-friendly web dashboard framework highlights the model's practical feasibility for delivering real-time diagnostic assistance in resource-constrained environments.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Puspita Kartikasari

Department of Statistics, Faculty of Science and Mathematics, Universitas Diponegoro

Prof. Soedarto, SH street, Tembalang, Semarang, Central Java 50275, Indonesia

Email: puspitakartikasari@live.undip.ac.id

1. INTRODUCTION

Pulmonary infections pose a severe global health burden, with pneumonia emerging as a critical respiratory condition. Although frequently viral in origin [1], it remains a leading cause of mortality among vulnerable populations, particularly children under five. According to the World Health Organization (WHO), pneumonia is a primary infection-driven cause of child mortality, disproportionately impacting developing nations [2]. In these resource-constrained regions, limited access to expert radiologists often results in delayed or inaccurate manual interpretations of diagnostic images, causing patients to miss the crucial window for effective therapeutic intervention. Consequently, developing an accessible, automated diagnostic tool is not merely a technological advancement, but an urgent public health necessity.

Pneumonia, stemming from bacterial or viral pathogens, necessitates distinct treatment pathways, making accurate diagnosis critical especially for vulnerable populations [3]. While chest X-rays remain the primary diagnostic modality due to their cost-efficiency and non-invasive nature, manual interpretation presents significant clinical challenges; suboptimal image quality, subtle symptom manifestation, and

observer fatigue can easily lead to diagnostic errors. To mitigate these discrepancies, artificial intelligence has emerged as a transformative tool, functioning as a reliable second opinion capable of discerning complex disease patterns directly from medical imagery [4], [5].

Consequently, this research establishes an automated diagnostic framework leveraging deep convolutional neural networks (DCNN) for robust feature extraction and pattern recognition [6], [7], designed with the scalability to potentially address other pulmonary conditions in the future. The selection of specific modeling techniques in this study is grounded in a critical evaluation of limitations found in recent literature. For instance, Jraibi and Zheng [8] successfully utilized depthwise separable convolution (DSC) within a hybrid architecture to enhance early Alzheimer's detection, the study remains severely limited by high computational complexity due to its integration with vision transformers, and it lacks a practical application platform for real-time clinical diagnosis. Similarly, Li *et al.* [9] applied transfer learning (TL) to address data scarcity in brain tumor segmentation; however, this research remains confined to theoretical predictive algorithms without being implemented into a web-based interface for direct clinical use. Furthermore, Tungoe and Doegar [10] proposed a deep learning ensemble method utilizing soft voting mechanisms, yet this approach demands substantial computational resources and has not been translated into a deployable clinical application. These gaps highlight the absolute necessity for a system that is not only accurate but also computationally efficient and ready for practical implementation.

Building on prior insights, this study develops a hybrid prediction model to distinguish between normal and pneumonia-infected lungs by integrating DSC and TL through an ensemble soft voting mechanism. By combining DSC's computational efficiency with the robust representational power of MobileNetV2, this architecture achieves rapid inference on standard clinical hardware without requiring expensive GPU clusters. This ensemble approach also improves adaptability to common medical imaging challenges, such as class imbalance and visual noise to maximize accuracy and stability, particle swarm optimization (PSO) dynamically optimizes the hyperparameters of both components, outperforming manual tuning method. Ultimately, rather than proposing a novel theoretical algorithm, this research makes a practical contribution by fusing these advanced techniques into a highly efficient diagnostic framework. Deployed as an interactive web application, this system bridges the gap between computational research and clinical practice, offering frontline healthcare workers an accessible decision-support tool ready for telemedicine integration.

2. METHOD

2.1. Data gathering

This study utilized a validated Kaggle dataset comprising 5,840 chest X-ray images, including 4,265 pneumonia cases and 1,575 normal cases. Figure 1 presents representative samples from the dataset, with Figure 1(a) showing a normal lung and Figure 1(b) depicting pneumonia. To ensure robust evaluation, the dataset was partitioned into training (70%), validation (20%), and testing (10%) subsets. Before model training, each image underwent a rigorous preprocessing pipeline involving resizing and pixel normalization. To address class imbalance and mitigate overfitting, data augmentation techniques such as random rotation, horizontal flipping, and zooming were applied, thereby encouraging the model to learn robust, position-invariant features rather than memorizing the training data.

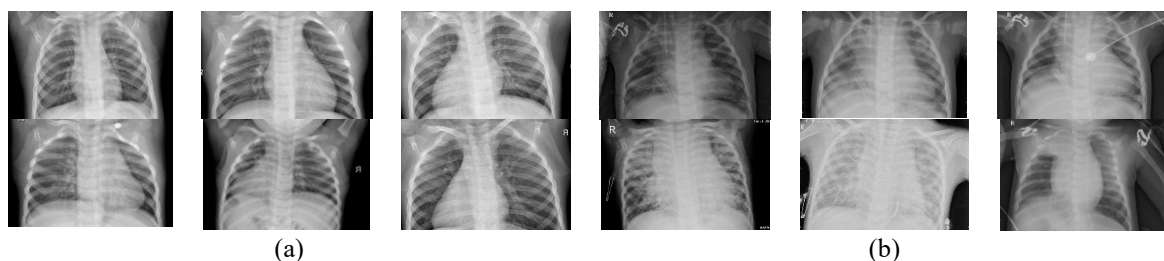


Figure 1. Sample image data for (a) normal lung and (b) pneumonia

2.2. Model architecture

The systematic workflow of this study is illustrated in Figure 2, which outlines the end-to-end experimental procedure to ensure methodological reproducibility. The research framework is structured into three sequential stages: data preparation, model training and optimization, and performance evaluation. This

logical organization directly mirrors the narrative flow of the subsequent results and discussion section, providing a clear and coherent guide for interpreting the findings.

The experimental workflow is structured into three primary phases. First, the raw dataset undergoes preprocessing, augmentation, and partitioning to ensure robust model training. Second, during the training and optimization phase, MobileNetV2 serves as the feature extraction backbone, while PSO is applied to enhance convergence and overall accuracy. Finally, the performance evaluation phase validates the framework's clinical efficacy in distinguishing normal from pneumonia cases using a confusion matrix alongside standard classification metrics.

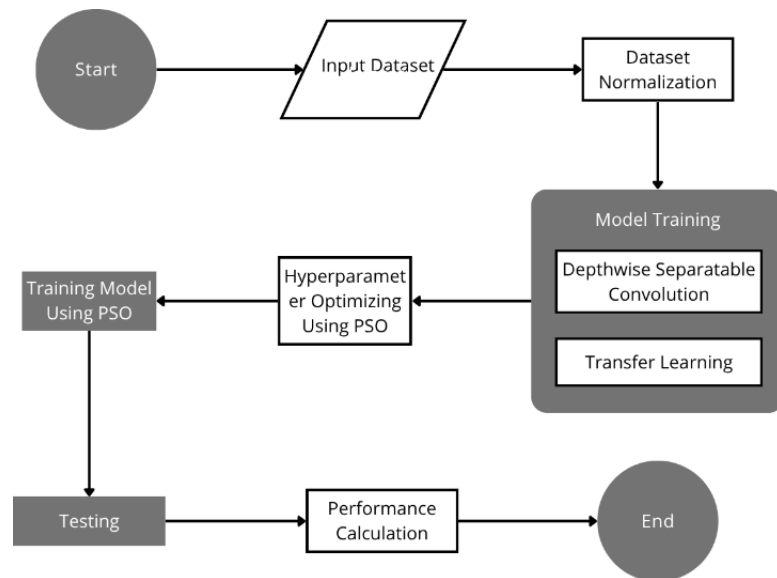


Figure 2. Systematic workflow of the proposed framework

2.2.1. Depthwise separable convolution

DSC is a technique strategically employed to significantly minimize the computational complexity and parameter footprint of the modeling process [11]. Functionally, this architectural method decouples the standard operation into two separate stages to integrate information between channels while simultaneously condensing feature dimensions at the output. This structural optimization preserves the method's high diagnostic performance while significantly alleviating the computational burden [12]. Furthermore, DSC minimizes the network's mathematical calculation process with only a negligible decrease in predictive accuracy, firmly ensuring that the execution is highly optimized [13].

For quantitative comparison, a standard convolution layer requires $D_K \cdot D_K \cdot M \cdot N$ parameters. In contrast, DSC significantly reduces this cost by splitting operation, requiring only $D_K \cdot D_K \cdot M + 1 \cdot 1 \cdot M \cdot N$ parameters [14]. This formula demonstrates how the 1×1 pointwise convolution drastically reduces computational overhead. Because high-resolution medical images often create severe memory bottlenecks with standard convolutions, this separable approach allows deeper networks to be trained efficiently on standard hardware. By isolating spatial filtering from cross-channel integration, the model retains strong representational power without the exponential parameter growth typical of dense layers. As illustrated in Figure 3, this specific DSC architecture effectively balances minimal computational complexity with robust feature extraction capabilities [15].

Figure 3 illustrates the sequential workflow of the DSC mechanism. Unlike standard convolutions that process spatial and channel dimensions concurrently, this architecture intentionally decouples the operation into two distinct stages. Starting with the input channels (M), the first stage depthwise convolution applies a single spatial filter independently to each channel, capturing spatial patterns without cross-channel mixing. Subsequently, the pointwise convolution stage employs a 1×1 kernel to compute a weighted sum across the depth dimension, acting as a linear projection that merges these independent channels. By dividing the feature extraction process into isolated spatial filtering and cross-channel integration, this approach

effectively generates the final output feature map while requiring a significantly lower parameter count than traditional convolutional methods.

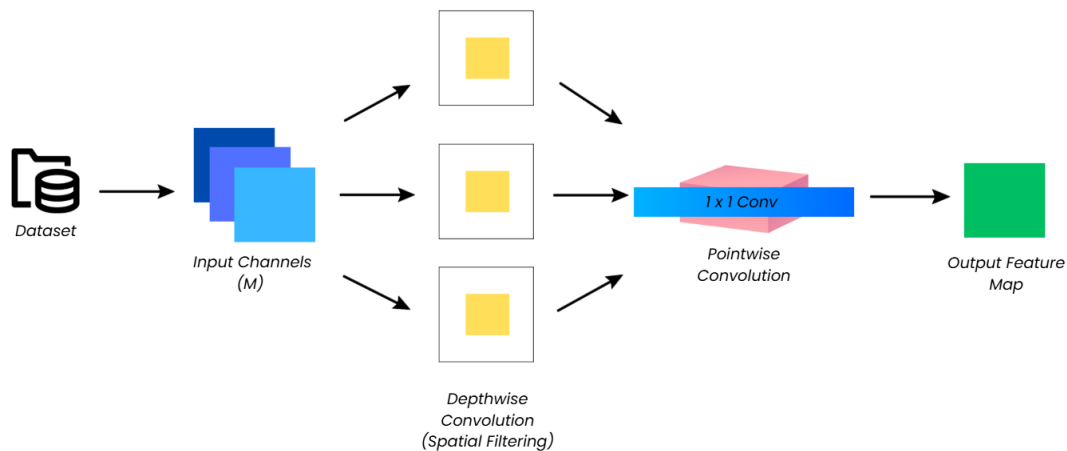


Figure 3. DSC architecture

2.2.2. Transfer learning with MobileNetV2

TL leverages pre-trained architectures as robust feature extractors for medical classification, eliminating the need to train from scratch and ensuring efficiency even with limited data and computing resources [16]. This study employs MobileNetV2 as the pre-trained backbone, a lightweight convolutional neural network specifically engineered for resource-constrained environments [17]. By achieving high classification accuracy with significantly fewer arithmetic operations than its predecessors [18]. This computational efficiency makes MobileNetV2 highly effective for rapid medical screening and diagnosis [19]. Its efficiency is driven by key architectural innovations—namely, the inverted residual structure and linear bottlenecks, which temporarily expand the feature space for complex non-linear transformations while preventing information loss in low-dimensional spaces. Ultimately, these features make MobileNetV2 uniquely suited to distill complex medical topologies into dense, informative vectors without generating overwhelming computational overhead.

Figure 4 illustrates the proposed architecture, which utilizes a pre-trained MobileNetV2 as the base feature extractor. To preserve the previously learned representations, the weights of this base model are frozen. A custom classification head is then appended to adapt the network for this specific task. Within this head, a global average pooling 2D layer first reduces the spatial dimensions of the feature maps, followed by a dropout layer to provide regularization and mitigate overfitting. Finally, a dense layer with a sigmoid activation function outputs a probability score, classifying the input as either normal or pneumonia.

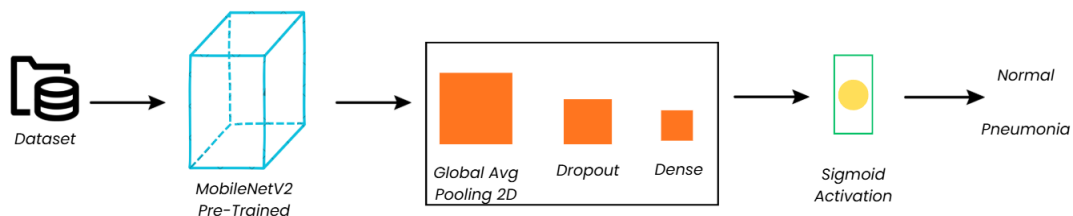


Figure 4. TL architecture

2.2.3. Ensemble soft voting

To enhance system robustness and mitigate overfitting, this study implements an ensemble soft voting mechanism [20]. Unlike hard voting which relies on rigid binary labels and discards crucial clinical nuance soft voting aggregates the continuous probability estimates from each classifier [21]. This approach is especially valuable in medical diagnostics, as it allows highly confident predictions to appropriately offset

uncertain ones in marginal cases, thereby improving overall predictive precision [22]. Ultimately, the final classification is determined by the weighted average of output probabilities across all models, ensuring that decisions are driven by the ensemble's collective confidence rather than a single isolated output [23].

2.3. Particle swarm optimization: conceptual foundations

PSO is a highly efficient stochastic optimization algorithm recognized for its rapid convergence [24]. Inspired by the natural foraging behaviors of bird flocks and fish schools, PSO represents candidate solutions as a collaborative swarm of particles. Unlike standard evolutionary algorithms that eliminate weaker individuals, PSO maintains the entire population throughout the search process. As these particles navigate a multidimensional search space, they continuously update their trajectories by combining their current velocity with their individual cognitive experience and the overall social intelligence of the swarm [25]. This dynamic movement is mathematically governed by iteratively updating the particle's velocity (v) and position (x), which are formally defined in (1) and (2).

$$v_i^{k+1} = w \cdot v_i^k + c_1 \cdot r(P_{best_i} - x_i^k) + c_2 \cdot r(G_{best_i} - x_i^k) \quad (1)$$

$$x_i^{k+1} = x_i^k + v_i^{k+1} \quad (2)$$

In (1) calculates the new velocity (v_i^{k+1}) necessary to update the particle's position in (2). This update mechanism is driven by three distinct forces. First, the inertia weight (w), typically ranging from 0.4 to 0, maintains previous momentum to balance global exploration and local exploitation; dynamically tuning this parameter allows for broad search in early epochs and precise fine-tuning in later stages, effectively preventing premature convergence. Second, the cognitive component represents the particle's memory of its personal best position (P_{best}), weighted by the coefficient c_1 . Third, the social component guides the particle toward the best position found by entire swarm (G_{best}), weighted by coefficient c_2 [26]. The learning factors c_1 and c_2 , generally range between 1.0 and 2.5, controlling the trade-off between independent exploration and swarm convergence. Maintaining this precise balance is critical for navigating the non-convex hyperparameter surfaces of deep learning frameworks, ensuring the algorithm escapes local optima to discover robust configurations that generalize well to unseen validation data.

2.4. Evaluation metrics

The careful selection of highly appropriate evaluation metrics plays a pivotal role in accurately assessing the real efficacy of a classification model. To fully ensure that the experimental results remain unbiased and accurately reflect the model's true diagnostic capabilities, the chosen metrics must align strictly with the characteristics of the dataset and the specific clinical objectives of the study. In the direct context of this research, standard evaluation metrics generally applicable to binary classification problems are utilized to validate the proposed method [27]. The foundational quantitative assessment of the proposed model is strictly derived from the confusion matrix. As clearly illustrated in Table 1, this matrix meticulously maps the relationship between the predicted classes and the actual ground truth medical labels. It provides a highly detailed breakdown of the classification performance by logically categorizing outcomes into four distinct components: true positive (TP), true negative (TN), false positive (FP), and false negative (FN).

Table 1. Confusion matrix

Actual	Predicted	
	Positive	Negative
Positive	True positive	False negative (type II error)
Negative	False positive (type II error)	True negative

Following the categorization of prediction outcomes within the confusion matrix, it is essential to derive quantitative metrics to accurately evaluate the model's performance. To rigorously validate clinical efficacy, this study avoids relying on a single indicator and instead adopts a multi-faceted approach utilizing accuracy, precision, recall, and the F1-score. This comprehensive combination provides a transparent assessment of the classifier's ability to reliably distinguish between positive and negative instances. The mathematical formulations for these evaluation metrics are defined as (3) to (6).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (6)$$

While accuracy is a fundamental metric, relying on it exclusively can be misleading when evaluating imbalanced datasets. In such scenarios, classifiers often bias toward the majority class, yielding a deceptively high accuracy while failing to identify minority instances [28]. To address this limitation, the F1-score calculated as the harmonic mean of precision and recall provides a more robust performance metric for imbalanced clinical data by penalizing severe disparities between false positives and false negatives. Consequently, evaluating precision, recall, and the F1-score alongside standard accuracy is crucial for a comprehensive and unbiased assessment of the model's true clinical efficacy.

3. RESULTS AND DISCUSSION

3.1. Data preparation

Before model development, the medical dataset was systematically partitioned to ensure a rigorous and unbiased evaluation. The chest X-ray images were divided into training, validation, and testing subsets using a strict 70:20:10 ratio. This deliberate separation is fundamental to the experimental design, as it provides a substantial data sample for model training while securely reserving dedicated subsets for hyperparameter tuning and final performance assessment [28]. The exact distribution of images across these categories is detailed in Table 2.

Table 2. The distribution of dataset

Category	Training	Validation	Testing
Normal	1,103	315	157
Pneumonia	2,986	853	426
Proportion	70%	20%	10%

As detailed in Table 2, the dataset was partitioned to support robust model development, allocating 70% for training to learn comprehensive feature representations, 20% for validation to monitor performance and mitigate overfitting, and the final 10% for testing to evaluate generalization on entirely unseen data. Following this split, a descriptive statistical analysis was conducted to examine the dataset's underlying numerical characteristics. Because machine learning algorithms process quantitative data rather than raw visuals, the X-ray images were converted into numerical arrays representing precise red, green, blue (RGB) pixel intensities, translating the visual information into a computable mathematical format. A comprehensive summary of these resulting numerical representations is provided in Table 3.

Table 3. The descriptive statistics of each category

Category	Minimum	Mean	Maximum	Median	Coefficient of variance (%)
Normal	0.000	122.404	254.913	131.575	49.89
Pneumonia	0.188	123.456	252.478	134.631	45.5

As detailed in Table 3, the descriptive statistics of pixel intensity distributions reveal subtle yet clinically relevant differences between the two classes. While both categories span nearly the entire 8-bit spectrum, the 'Normal' class exhibits a slightly wider range (0.000 to 254.913) compared to the 'Pneumonia' class (0.188 to 252.478). Additionally, pneumonia-affected X-rays display a marginally brighter central tendency, evidenced by an elevated mean (123.456) and median (134.631) relative to the normal baseline (mean: 122.404, median: 131.575). Analysis of data dispersion further highlights these distinctions; the 'Normal' category shows greater overall variability with a coefficient of variance of 49.89%, whereas the 'Pneumonia' category's lower variance of 45.5% indicates tighter clustering around the mean. This increased pixel homogeneity in pneumonia cases strongly aligns with the visual pathology of lung consolidation, where fluid accumulation generates uniformly opaque regions that convolutional filters are particularly adept at detecting.

3.2. Data preprocessing

Following the initial assessment of raw pixel intensities, data normalization serves as a critical preprocessing step to accelerate neural network convergence and ensure compatibility with the pre-trained architecture. By rescaling the original 8-bit pixel values into a standardized range of -1 to 1, this procedure preserves the intrinsic structural patterns of the X-ray images while eliminating scale disparity. This essential transformation guarantees that the model evaluates all input features equally, preventing high-intensity pixels from disproportionately biasing the learning process [19]. The quantitative impact of this normalization step is summarized in Table 4.

Table 4. The descriptive statistics of each category after normalization

Category	Minimum	Mean	Maximum	Median
Normal	-1.000	-0.040	1.000	0.032
Pneumonia	-1.000	-0.020	1.000	0.069

The normalization process successfully remapped the dataset into a target interval centered around zero, a transformation critical for the tanh and sigmoid activation functions in the TL model to maintain steep gradients during backpropagation and prevent the vanishing gradient problem [7]. Despite this significant numerical compression, the original distribution characteristics were well-preserved. Specifically, the ‘pneumonia’ category maintained slightly higher central tendency values with an adjusted mean of -0.020 and a median of 0.069 compared to the ‘normal’ category’s mean of -0.040 and median of 0.032, ensuring that the fundamental diagnostic distinctions between the classes remain intact. To provide a qualitative perspective on this quantitative shift, Figure 5 presents effect of normalization on a representative normal lung X-ray image. Figure 5(a) shows the raw image before normalization, while Figure 5(b) presents the corresponding normalized image. Although the clinical and pathological features remain visually consistent to human observers, this precise normalization is essential for ensuring the computational efficiency and stability of the deep learning framework.

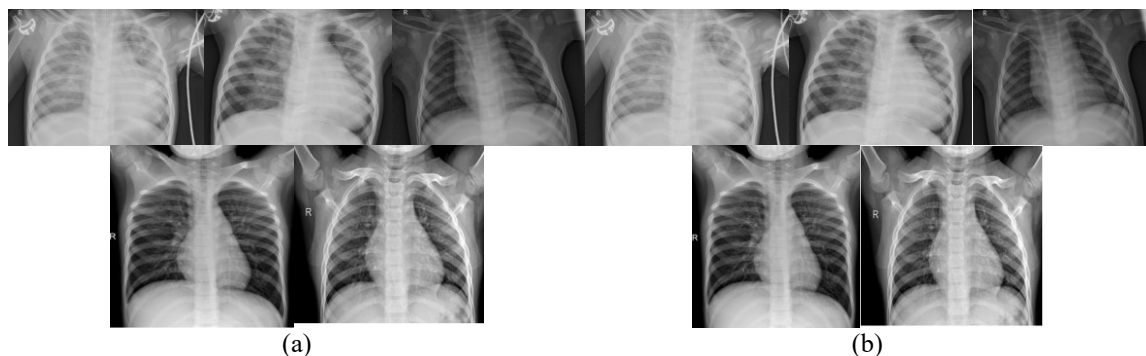


Figure 5. Normal lung image data (a) before normalization and (b) after normalization

3.3. Modelling

In this study, two distinct deep learning architectures were developed and evaluated: the custom DSC model and the TL model based on MobileNetV2. To ensure optimal performance, the hyperparameters for both models were fine-tuned using PSO [25]. The specific structural configuration of each model—including the arrangement of layers, output shapes, and the number of trainable parameters—is detailed below to provide a transparent view of the computational complexity involved [13]. The comprehensive architectural specifications for the best-performing models are presented in Table 5.

As detailed in Table 5, the custom DSC model generates 3,218,300 trainable parameters, predominantly concentrated within the fully connected layers due to the structural bottleneck of unrolling spatial data. Conversely, the TL model leverages a pre-trained MobileNetV2 base to achieve a more compact architecture of just 2,586,177 parameters. By employing a PSO-optimized classification layer of 256 neurons, the TL framework successfully captures essential feature complexities while avoiding the dense parameter overhead inherent to the custom model. Consequently, this computationally lighter TL approach not only conserves valuable training resources but also actively mitigates the risk of overfitting.

Table 5. The best architecture

Model	Layer	Output shape	Param #
DSC	Input layer	(None, 224, 224, 3)	0
	Separable_conv2d	(None, 224, 224, 32)	123
	Separable_conv2d_1	(None, 112, 112, 32)	1,312
	Separable_conv2d_2	(None, 56, 56, 32)	1,312
	Separable_conv2d_3	(None, 28, 28, 32)	1,312
	Separable_conv2d_4	(None, 14, 14, 32)	1,312
	Flatten	(None, 6272)	0
	Dropout	(None, 512)	0
	Dense	(None, 512)	3,211,776
	Dense_1	(None, 1)	513
Total params			3,218,300
TL	Input layer	(None, 224, 224, 3)	0
	MobileNetV2	(None, 7, 7, 1280)	2,257,984
	Dropout	(None, 256)	0
	Dense	(None, 128)	327,936
	Dense_1	(None, 1)	256
Total params			2,586,177

3.4. Soft voting

To further improve predictive performance and robustness, this study introduces a soft voting ensemble strategy. In contrast to hard voting, which relies on a simple majority rule, soft voting aggregates the predicted probabilities from both the DSC and TL models [10]. The final prediction is then derived by computing a weighted average of these confidence scores. By relying on continuous probabilities rather than flat labels, this method naturally penalizes false certainty and minimizes individual model biases, leading to much more stable and reliable performance on unseen data [21]. Ultimately, this approach seamlessly combines the TL's spatial analysis with the DSC's structural hierarchy, fusing their complementary strengths into a single, highly cohesive diagnostic tool.

3.5. Particle swarm optimization: hyperparameter and ensemble weight tuning

PSO plays a critical dual role in this research. Primarily, it serves as a global optimization algorithm to automate the selection of optimal hyperparameters for the individual deep learning models [25]. By navigating the search space effectively, PSO identifies the best configuration for parameters such as learning rate, dropout rate, and activation functions. The specific hyperparameter values derived from this optimization process for both the DSC and TL models are detailed in Table 6.

Table 6. Hyperparameter PSO

Model	Hyperparameter	Output
DSC	Dropout rate	0.1000
	Learning rate	0.0010
	Activation function	Sigmoid
	Optimizer	SGD
	Number of conv layers	5
	Filters value	32
	Dense unit	512
TL	Dropout rate	0.1000
	Learning rate	0.0010
	Activation function	Tanh
	Optimizer	Adam
	Dense units	256

Beyond merely tuning hyperparameters, PSO is specifically and strategically applied to perfectly optimize the specific voting weights within the soft voting ensemble [22]. Instead of blindly assigning equal importance (a static 50:50 split) to both distinct models, PSO dynamically and intelligently searches for the optimal weight combination that mathematically maximizes the final classification performance. This rigorous process absolutely ensures that the model possessing higher predictive confidence inherently contributes far more significantly to the final diagnostic decision. To evaluate the true effectiveness of this exact optimization strategy, the performance of the full ensemble model is benchmarked directly against the individual models. The resulting F1-score comparison is visually presented in Figure 6.

As illustrated in Figure 6, the TL model demonstrates incredibly strong standalone performance with a high F1-score of 0.9198, significantly outperforming the custom DSC model, which only attained an F1-score of 0.8884. However, the exact implementation of the ensemble soft voting coupled with

PSO-optimized weights yield the absolute highest result, achieving an impressive F1-score of 0.9214. This concrete improvement absolutely confirms that the ensemble strategy highly successfully synergizes the distinct strengths of both architectures, ultimately resulting in a far more reliable prediction model than either of the individual classifiers operating alone.

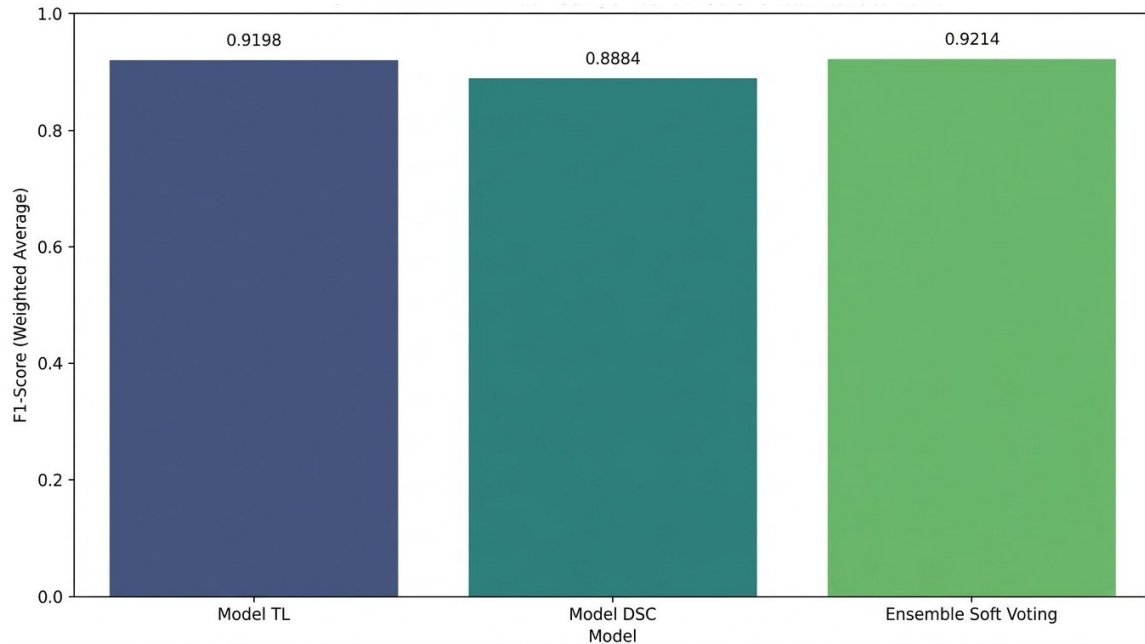


Figure 6. F1-score comparison chart

3.6. Discussion

As a final, definitive evaluation, a highly comprehensive comparison of the classification performance achieved by the proposed models was meticulously conducted. The evaluation closely analyzed confusion matrices, loss values, and accuracy metrics consistently throughout the entire training and validation phases. The detailed performance outcomes for PSO-DSC model are illustrated in Figures 7 and 8.

Figure 6 shows the F1-score comparison chart of the PSO-DSC, PSO-TL, and PSO-ensemble models. Figure 7 shows the confusion matrix of the PSO-DSC model, detailing classification rates, alongside the loss curve in Figure 8(a) and the accuracy curve in Figure 8(b) that track the learning process. Although the stochastic nature of the SGD optimizer causes some early, jagged fluctuations in the validation loss, PSO effectively stabilizes the model in later epochs by fine-tuning dropout and learning rates [25]. Building on this, the study then evaluates the TL architecture, which is similarly optimized with PSO. As shown in Figure 8, the TL model is assessed using the exact same metrics confusion matrix, loss trajectory, and accuracy curves to map out its learning dynamics and overall generalization potential [18].

Figure 9 illustrates the performance of the PSO-TL model. Figure 9(a) shows the training and validation accuracy, while Figure 9(b) shows the training and validation loss, which track training stability. These metrics clearly show that MobileNetV2, powered by the Adam optimizer and pre-trained ImageNet weights, achieves significantly faster convergence than the standalone DSC model. To maximize these individual strengths, this study combined the PSO-DSC and PSO-TL architectures using a soft voting ensemble technique. Figure 10 presents the confusion matrix, offering a comprehensive overview of the ensemble model's performance on the validation set. It details the distribution of accurate and inaccurate predictions, specifically highlighting the true positives and true negatives derived from the combined probabilistic outputs of the base models. This visualization confirms the ensemble's ability to effectively distinguish between normal and pneumonia cases, demonstrating the practical advantage of the soft voting mechanism in reducing misclassification rates. To consolidate these findings, the predictive performance of the DSC and TL models, which was initially evaluated individually, is now benchmarked against the ensemble approach. The detailed performance metrics for each architecture, including the Soft Voting ensemble optimized via PSO, are comprehensively summarized in Table 7.

		<i>Predicted</i>	
		<i>Positive</i>	<i>Negative</i>
<i>Actual</i>	<i>Positive</i>	141	16
	<i>Negative</i>	51	375

Figure 7. Confusion matrix of the PSO-DSC model

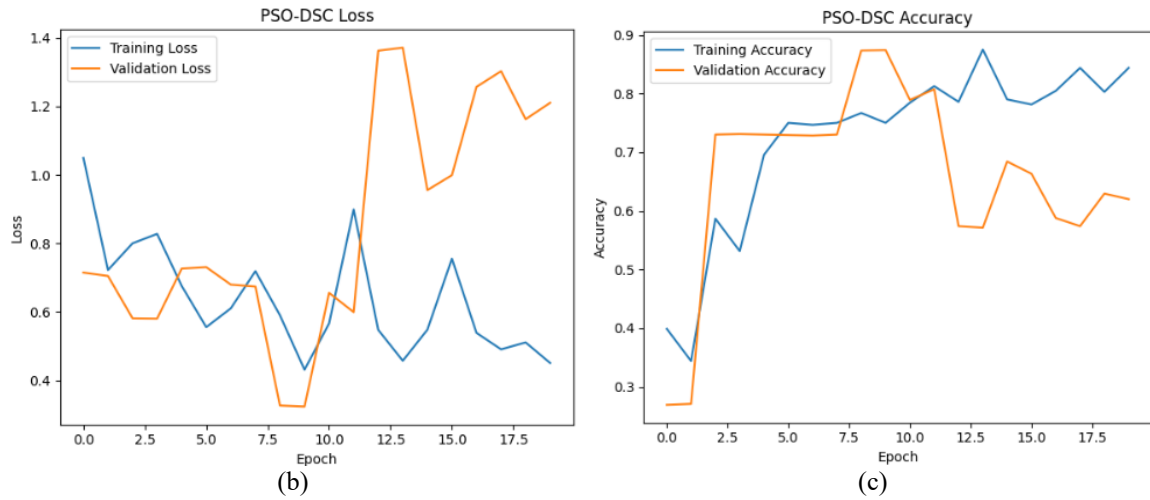


Figure 8. Performance evaluation of the PSO-DSC model for (a) loss and (b) accuracy



Figure 9. Performance evaluation of the PSO-TL model for (a) accuracy and (b) loss

		<i>Predicted</i>	
		<i>Positive</i>	<i>Negative</i>
<i>Actual</i>	<i>Positive</i>	136	21
	<i>Negative</i>	26	400

Figure 10. Confusion matrix of the PSO-TL model

Table 7. Prediction result

Evaluation	DSC (%)	TL (%)	Soft voting (%)
Accuracy	88.51	91.93	92.11
Precision	95.91	95.01	95.02
Recall	88.02	93.89	94.13

The soft voting ensemble model, optimized with PSO, achieved superior diagnostic performance with 92.11% accuracy, 95.02% precision, and 94.13% recall by effectively resolving difficult edge cases that standalone architectures often miss. In a practical clinical setting, this high recall crucially minimizes the risk of dangerously delayed treatments by preventing false negatives, while the strong precision simultaneously spares healthy patients from unnecessary medical interventions and wasted hospital resources. Beyond its binary classification capabilities, the model demonstrated remarkable statistical certainty and stability in its predictions, evidenced by exceptionally low error rates mean squared error (MSE): 0.0626, root mean square error (RMSE): 0.2501, mean absolute error (MAE): 0.1213 and a strong R-squared value of 68.21%. Leveraging this balance of high diagnostic accuracy and robust probabilistic calibration, the framework was successfully deployed into a lightweight, web-based computer-aided diagnosis (CAD) system that provides medical professionals with a real-time, highly accessible tool for fast-paced environments. The resulting web-based interface for pneumonia prediction is illustrated in Figure 11, demonstrating the practical implementation of the proposed CAD framework.

However, although the highly optimized model has achieved remarkable accuracy and incredibly good predictive performance, this specific study acknowledges notable limitations regarding its absolute generalization ability. The deep learning model was trained and rigorously evaluated using a single, static public dataset obtained exclusively from Kaggle. Consequently, the flawless consistency of the model's predictive performance on raw images stemming from entirely different demographic backgrounds, scanned with varying X-ray machine calibrations, or operating in vastly different real-world clinical settings has not yet been conclusively verified. Heavily addressing these critical limitations related to comprehensive data diversity is absolutely crucial before attempting full-scale, worldwide clinical implementation.

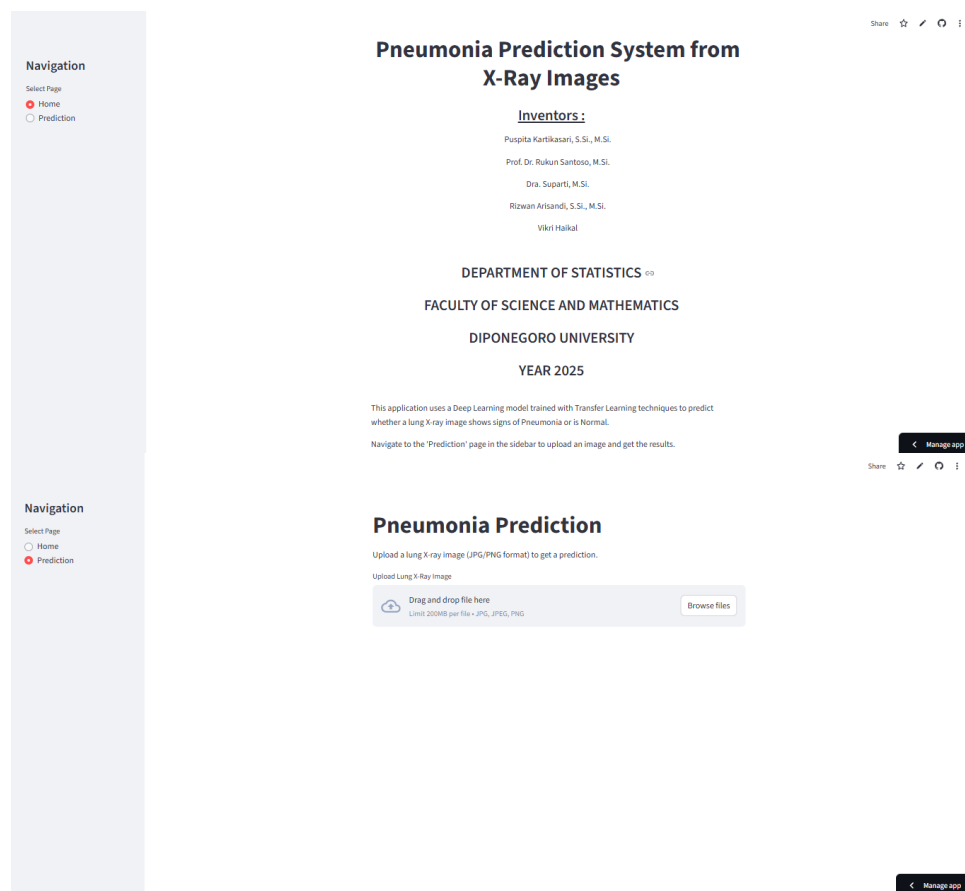


Figure 11. Pneumonia prediction website display

4. CONCLUSION

This research developed a reliable pneumonia detection framework by combining DSC and TL into a soft voting ensemble, with PSO employed to fine-tune both the models and their voting weights. The optimized ensemble significantly outperformed standalone approaches, achieving 92.11% accuracy with strong statistical stability, as evidenced by low error rates (RMSE: 0.2501, MSE: 0.0626, MAE: 0.1213) and a robust R-squared value of 68.21%. To demonstrate practical utility for early clinical intervention, the system was integrated into an interactive web dashboard, enabling frontline medical staff to access sophisticated models such as MobileNetV2 with ease. Nevertheless, given that the study currently relies on a single dataset, future work should prioritize validation across diverse, multi-center data and investigate edge-computing deployment to ensure reliable remote screening.

FUNDING INFORMATION

The researcher would like to thank the Research and Development Institute, Universitas Diponegoro, Semarang, Indonesia, under highly reputable international research grant number 222-680/UN7.D2/PP/IV/2025, for the research funding support provided, which was very helpful in the successful completion of this research.

AUTHOR CONTRIBUTIONS STATEMENT

The journal applies the Contributor Role Taxonomy (CRediT) to identify each author’s role, minimize authorship conflicts, and support collaboration among researchers.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Puspita Kartikasari	✓	✓	✓	✓	✓	✓			✓	✓		✓	✓	✓
Rukun Santoso	✓				✓		✓		✓	✓				
Suparti	✓							✓	✓	✓				
Rizwan Arisandi	✓		✓							✓	✓	✓		
Vikri Haikal	✓							✓		✓				

C : Conceptualization	I : Investigation	Vi : Visualization
M : Methodology	R : Resources	Su : Supervision
So : Software	D : Data Curation	P : Project administration
Va : Validation	O : Writing - Original Draft	Fu : Funding acquisition
Fo : Formal analysis	E : Writing - Review & Editing	

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

DATA AVAILABILITY

The data on which the findings of this study are based are freely accessible through <https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia>.

REFERENCES

[1] J. Zhang *et al.*, “Viral pneumonia screening on chest X-rays using confidence-aware anomaly detection,” *IEEE Transactions on Medical Imaging*, vol. 40, no. 3, pp. 879–890, Mar. 2021, doi: 10.1109/TMI.2020.3040950.

[2] World Health Organization, “Pneumonia in children,” *who.int*, Geneva, Switzerland: World Health Organization, Nov. 2021. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/pneumonia>

[3] X. Lan, Y. Zhang, W. Yuan, F. Shi, and W. Guo, “Image-based deep learning in diagnosing mycoplasma pneumonia on pediatric chest X-rays,” *BMC Pediatrics*, vol. 24, no. 1, Nov. 2024, doi: 10.1186/s12887-024-05204-0.

[4] J. M. Kimeu, M. Kisangiri, H. Mbelwa, and J. Leo, “Deep learning-based mobile application for the enhancement of pneumonia medical imaging analysis: a case-study of West-Meru Hospital,” *Informatics in Medicine Unlocked*, vol. 50, 2024, doi: 10.1016/j.imu.2024.101582.

[5] M. Kaya, “Feature fusion-based ensemble CNN learning optimization for automated detection of pediatric pneumonia,” *Biomedical Signal Processing and Control*, vol. 87, Jan. 2024, doi: 10.1016/j.bspc.2023.105472.

- [6] M. E. Rayed, S. M. S. Islam, S. I. Niha, J. R. Jim, M. M. Kabir, and M. F. Mridha, "Deep learning for medical image segmentation: State-of-the-art advancements and challenges," *Informatics in Medicine Unlocked*, vol. 47, 2024, doi: 10.1016/j.imu.2024.101504.
- [7] S. Iqbal, A. N. Qureshi, J. Li, and T. Mahmood, "On the analyses of medical images using traditional machine learning techniques and convolutional neural networks," *Archives of Computational Methods in Engineering*, vol. 30, no. 5, pp. 3173–3233, Jun. 2023, doi: 10.1007/s11831-023-09899-9.
- [8] N.-E. H. Jraibi and H. Zeng, "DSCViTANet: a hybrid depthwise separable convolution and vision transformer for early alzheimer's classification," in *2025 10th International Conference on Image, Vision and Computing (ICIVC)*, IEEE, Jul. 2025, pp. 106–111. doi: 10.1109/ICIVC66358.2025.11200357.
- [9] Y. Li, S. Jiang, Z. Yang, L. Wang, S. Wang, and Z. Zhou, "Brain tumor segmentation via cross-modality semi-supervised transfer learning with 3D MRI diffusion model synthetic ultrasound," *Information Fusion*, vol. 127, Mar. 2026, doi: 10.1016/j.inffus.2025.103757.
- [10] N. Y. Tungoe and A. Doegar, "Classification of brain tumor using ensemble of deep learning models," in *International Conference on Information Systems and Management Science*, 2025, pp. 710–719. doi: 10.1007/978-3-031-95017-9_62.
- [11] K. N. Ramadhani, R. Munir, and N. P. Utama, "Improving video vision transformer for deepfake video detection using facial landmark, depthwise separable convolution and self-attention," *IEEE Access*, vol. 12, pp. 8932–8939, 2024, doi: 10.1109/ACCESS.2024.3352890.
- [12] R. Qi, R.-S. Jia, Q.-C. Mao, H.-M. Sun, and L.-Q. Zuo, "Face detection method based on cascaded convolutional networks," *IEEE Access*, vol. 7, pp. 110740–110748, 2019, doi: 10.1109/ACCESS.2019.2934563.
- [13] B. Liu, D. Zou, L. Feng, S. Feng, P. Fu, and J. Li, "An FPGA-based CNN accelerator integrating depthwise separable convolution," *Electronics*, vol. 8, no. 3, Mar. 2019, doi: 10.3390/electronics8030281.
- [14] Y. Zhao, L. Bai, Y. Lyu, and X. Huang, "Camera-based blind spot detection with a general purpose lightweight neural network," *Electronics*, vol. 8, no. 2, Feb. 2019, doi: 10.3390/electronics8020233.
- [15] M. A. Hossain and M. S. A. Sajib, "Classification of image using convolutional neural network (CNN)," *Global Journal of Computer Science and Technology*, pp. 13–18, 2019, doi: 10.34257/gjcsdvol19is2pg13.
- [16] M. Shakya, R. Patel, and S. Joshi, "A comprehensive analysis of deep learning and transfer learning techniques for skin cancer classification," *Scientific Reports*, vol. 15, no. 1, Feb. 2025, doi: 10.1038/s41598-024-82241-w.
- [17] D. Panchal et al., "MobileNet-based transfer learning: a novel approach for improved Alzheimer's disease classification from brain imaging," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 3, pp. 2777–2783, 2024.
- [18] H. Bichri, A. Chergui, and M. Hain, "Image classification with transfer learning using a custom dataset: comparative study," *Procedia Computer Science*, vol. 220, pp. 48–54, 2023, doi: 10.1016/j.procs.2023.03.009.
- [19] T. B.-S. Nguyen et al., "Applying machine learning with MobileNetV2 model for rapid screening of vaginal discharge samples in vaginitis diagnosis," *Scientific Reports*, vol. 15, no. 1, May 2025, doi: 10.1038/s41598-025-04626-9.
- [20] Y. Liao, Y. Gao, F. Wang, L. Zhang, Z. Xu, and Y. Wu, "Emotion recognition with multiple physiological parameters based on ensemble learning," *Scientific Reports*, vol. 15, no. 1, Jun. 2025, doi: 10.1038/s41598-025-96616-0.
- [21] I. Chhillar and A. Singh, "An improved soft voting-based machine learning technique to detect breast cancer utilizing effective feature selection and SMOTE-ENN class balancing," *Discover Artificial Intelligence*, vol. 5, no. 1, Jan. 2025, doi: 10.1007/s44163-025-00224-w.
- [22] S. Benghazouani, S. Nouh, and A. Zakrani, "Optimizing breast cancer diagnosis: Harnessing the power of nature-inspired metaheuristics for feature selection with soft voting classifiers," *International Journal of Cognitive Computing in Engineering*, vol. 6, pp. 1–20, Dec. 2025, doi: 10.1016/j.ijcce.2024.09.005.
- [23] D. Asif, M. S. Arif, and A. Mukheimer, "A data-driven approach with explainable artificial intelligence for customer churn prediction in the telecommunications industry," *Results in Engineering*, vol. 26, Jun. 2025, doi: 10.1016/j.rineng.2025.104629.
- [24] F. D. Wihartiko, H. Wijayanti, and F. Virgantari, "Performance comparison of genetic algorithms and particle swarm optimization for model integer programming bus timetabling problem," *IOP Conference Series: Materials Science and Engineering*, vol. 332, Mar. 2018, doi: 10.1088/1757-899X/332/1/012020.
- [25] D. P. Rini, T. K. Sari, W. K. Sari, and N. Yusliani, "Hyperparameter optimization of convolutional neural network using particle swarm optimization for emotion recognition," *IAES International Journal of Artificial Intelligence*, vol. 14, no. 1, pp. 547–560, Feb. 2025, doi: 10.11591/ijai.v14.i1.pp547-560.
- [26] Y. Shi and R. Eberhart, "A modified particle swarm optimizer," in *1998 IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98TH8360)*, IEEE, 1998, pp. 69–73. doi: 10.1109/ICEC.1998.699146.
- [27] J. S. Yang, Z. Zhang, and X. Xiao, "Adaptive reinforcement ensemble transfer learning for robust lung tumor detection," *Biomedical Signal Processing and Control*, vol. 114, Apr. 2026, doi: 10.1016/j.bspc.2025.109247.
- [28] F. Thabtah, S. Hammoud, F. Kamalov, and A. Gonsalves, "Data imbalance in classification: experimental evaluation," *Information Sciences*, vol. 513, pp. 429–441, 2020, doi: 10.1016/j.ins.2019.11.004.

BIOGRAPHIES OF AUTHORS



Puspita Kartikasari    holds her master's degree in Statistics from Institut Teknologi Sepuluh Nopember Surabaya in 2015. She also earned her bachelor's degree in Statistics from Institut Teknologi Sepuluh Nopember Surabaya in 2013. Currently, she serves as an associate lecturer in the Department of Statistics, Faculty of Science and Mathematics, Universitas Diponegoro, Semarang, Indonesia. She is in the Computation and Data Science Laboratory. Her research includes machine learning, computation, classification techniques in healthcare, time series analysis, and simulation. She has published more than 10 papers in national journals and international conferences. She can be contacted at email: puspitakartikasari@live.undip.ac.id.



Rukun Santoso    has been a lecturer and researcher at Universitas Diponegoro since January 1, 1992, after completing his undergraduate education at the Department of Mathematics, Faculty of Engineering, Universitas Diponegoro. He is a master's degree in Mathematics with a concentration in Statistics was obtained at Universitas Gadjah Mada in 2000. He is a doctoral degree in Mathematics with a concentration in Statistics was obtained at Universitas Gadjah Mada in 2016. The functional position of professor in the field of Statistical Computing is achieved since December 1, 2024. His research includes wavelets for statistics and machine learning. He can be contacted at email: rukunsantoso@lecturer.undip.ac.id.



Suparti    is a lecturer in the Department of Statistics, Faculty of Science and Mathematics, Universitas Diponegoro. She graduated with a Bachelor of Mathematics from the Faculty of Mathematics and Natural Sciences, Universitas Diponegoro in 1989 and a Master of Mathematics from Universitas Gadjah Mada, Indonesia in 1999, specializing in Statistics. Her current position is associate professor. She teaches nonparametric regression analysis, mathematical statistics, and wavelet methods. Her research interests include nonparametric regression analysis and wavelet methods. She can be contacted at email: suparti@lecturer.undip.ac.id.



Rizwan Arisandi    holds a Bachelor of Mathematics (S.Si) Master of Science (M.Si) in Statistics from the Institut Teknologi Sepuluh Nopember, Indonesia. He is currently lecturing with the Department of Computer Science at Universitas Bina Nusantara, Indonesia. His research areas of interest include the application of regression and classification methods in healthcare to predict disease risk and accurately identify patient conditions. By utilizing clinical data and machine learning algorithms. He can be contacted at email: rizwan.arisandi@binus.ac.id.



Vikri Haikal    is an active student in the Department of Statistics, Faculty of Science and Mathematics, Universitas Diponegoro. He is currently actively involved in academic research and independent studies focusing on financial modeling, risk analysis, and portfolio optimization. His areas of interest include machine learning applications, time series forecasting, and investment portfolio strategies. He has also explored deep learning applications in medical imaging and decision support systems. He can be contacted at email: haikalvikri360@gmail.com.