Hybrid feature fusion from multiple CNN models with bayesian-optimized machine learning classifiers

Dewi Rismawati, Sugiyarto Surono, Aris Thobirin
Department of Mathematics, Faculty of Science and Applied Technology, Ahmad Dahlan University, Yogyakarta, Indonesia

Article Info

Article history:

Received Apr 11, 2025
Revised Jun 21, 2025
Accepted Jul 4, 2025

Keywords:

Bayesian optimization
Big data
Convolutional neural network
Fusing feature
Machine learning

ABSTRACT

Information technology advancements have created big data, necessitating efficient techniques to retrieve helpful information. With its capacity to recognize and categorize patterns in data, especially the growing amount of picture data, deep learning is becoming a viable option. This research aims to develop a medical image classification model using chest X-Ray with four classes, namely Covid-19, Pneumonia, Tuberculosis, and Normal. The proposed method combines the advantages of deep learning and machine learning. Three pre-trained CNN models, VGG16, DenseNet201, and InceptionV3, extract features from images. The features generated from each model are fused to enhance the relevant information. Furthermore, principal component analysis (PCA) was applied to reduce the dimensionality of the features, and Bayesian optimization was used to optimize the hyperparameters of the machine learning algorithms support vector machine (SVM), decision tree (DT), and k-nearest neighbors (k-NN). The resulting classification model was evaluated based on accuracy, precision, recall, and F1-score. The results showed that FF-SVM, which is the proposed model, achieved an accuracy of 98.79% with precision, recall, and F1-score of 98.85%, 98.82%, and 98.84%, respectively. In conclusion, fusing feature extraction from multiple CNN models improved the classification accuracy of each machine-learning model. It provided reliable and accurate predictions for lung image diagnosis using chest X-Ray.

Corresponding Author:

Sugiyarto Surono
Department of Mathematics, Faculty of Science and Applied Technology, Ahmad Dahlan University
South Ring Road Street, Tamanan, Banguntapan, Bantul, Special Region of Yogyakarta 55191, Indonesia
Email: sugiyarto@math.uad.ac.id

1. INTRODUCTION

Pulmonary diseases such as COVID-19, pneumonia, and tuberculosis are global health problems that significantly impact morbidity and mortality rates [1]. Reducing the spread and improving patient prognosis requires an early and precise diagnosis. However, there are drawbacks to manually interpreting chest medical imaging, such as CT scans and radiographs, including the need for radiologists' knowledge, the possibility of human mistakes, and a lack of resources in places with low healthcare access [2].

Since the mid-20th century, the world has undergone profound changes thanks to rapid advances in information technology, particularly computer technology. Big Data, the result of these developments in the information age, is now an important force driving the transformation of various sectors, including healthcare. As the number and complexity of medical images increase in the Big Data era, conventional approaches to data processing become inadequate to meet the need for accurate and efficient analysis [3]. Recent developments, especially in deep learning (DL), have enabled the identifying, measuring, and classifying of patterns in image data [4]. DL is an algorithm inspired by how the human brain works, where artificial neural networks function like the network of neurons in our brain to process information and learn from experience [5], [6]. One of the most popular DL architectures in image data processing is convolutional neural networks (CNN). CNN is specifically designed to extract features representing various contexts of image data without feature engineering through convolution layers [7]. Several advanced CNN architectures were developed and applied in image data classification and pattern recognition [8].

Previous research by Gupta and Chawla [9] evaluated the efficiency of several pre-trained CNN models, including VGG16, VGG19, Xception, and ResNet50, along with support vector machine (SVM) and logistic regression (LR) for breast cancer classification using histopathology images. The results showed that ResNet50+LR achieved the best accuracy (93.27%), outperforming the other models. Research by Aslan *et al*. [10] developed a classification method for COVID-19 diagnosis using chest computed tomography (CT) images. They utilized several advanced CNN architectures (AlexNet, ResNet18, ResNet50, Inceptionv3, Densenet201, Inceptionresnetv2, MobileNetv2, GoogleNet) that had been pre-trained to extract features and classify them using several machine learning algorithms, where the DenseNet-SVM architecture gave the highest accuracy of 96.29%. Research by Biswas and Islam [11] classifies brain tumors through a hybrid model based on deep CNN (DCNN) and SVM. CNN+SVM obtained 96.0% accuracy, 98.0% specificity, and 95.71% sensitivity, higher than other transfer learning models (AlexNet, GoogLeNet, and VGG16).

Previous research has focused on utilizing individual CNN architectures for feature extraction, which are then combined with classical machine learning algorithms for classification. Therefore, this study proposes a feature fusion approach by fusing features from several pre-trained CNN architectures, namely VGG16, DenseNet201, and InceptionV3. This approach aims to utilize the advantages of each architecture in extracting different visual representations from medical images, resulting in more informative and rich features [12]. The principal component analysis (PCA) dimensionality reduction technique is then used to reduce the dimensionality of the data and transform the features into a more compact subspace while retaining important information [13].

Furthermore, classification is performed using several machine learning algorithms, namely SVM, decision tree (DT), and K-nearest neighbors (k-NN), which are known for their respective advantages. SVM offers significant advantages with its ability to process high-dimensional data and its computational efficiency [14]. DT is commonly applied because it is easy to interpret, trains rapidly, and can manage both numerical and categorical variables [15]. The benefits of k-NN approaches are straightforward to comprehend and execute [16]. Hyperparameter adjustment is an important component in training supervised and unsupervised ML models. Therefore, ML methods must be configured before training to get maximum results. This is because configuration variables affect model performance and accuracy [17]. Bayesian optimization (BO) is selected for hyperparameter tuning owing to its consistent advantage in reducing computational time relative to both Grid and random search methods [18].

Several studies evaluated the feature fusion of multiple pre-trained CNNs before applying classical ML classifiers. Alzahem [19] used Dempster-Shafer fusion on multiple CNNs but relied on ensemble theory rather than classical ML. Zhang *et al*. [20] combined CNN features but focused only on optimized CNNs without exploring the fusion of classical classifiers with Bayesian tuning. Overall, feature fusion combined with Bayesian-optimized classical machine learning classifiers remains unexplored. The main contributions of this research are summarized as follows:

– Multi-CNN feature fusion: features are extracted from three pre-trained CNN architectures-VGG16, DenseNet201, and InceptionV3-and combined using a feature fusion technique to enhance feature richness and diversity.
– Dimensionality reduction using PCA: PCA is applied to the fused features to reduce dimensionality, thereby improving computational efficiency and minimizing overfitting.
– Classification with optimized machine learning models: the dimensionality-reduced features are categorized into four types of lung diseases using optimized versions of SVM, DT, and k-NN algorithms, where the optimal hyperparameters are determined through BO.
– Clinically applicable classification pipeline: the proposed approach aims to develop a highly accurate and efficient classification system that can be integrated into clinical decision support systems for radiological diagnosis.

2. METHOD

This research uses a hybrid methodology integrating DL and machine learning to analyze lung image data. The stages of this research can be seen in Figure 1. It starts with image input, which is then reduced and normalized. Next, feature extraction is performed using three CNN models, namely VGG16, DenseNet201, and InceptionV3.

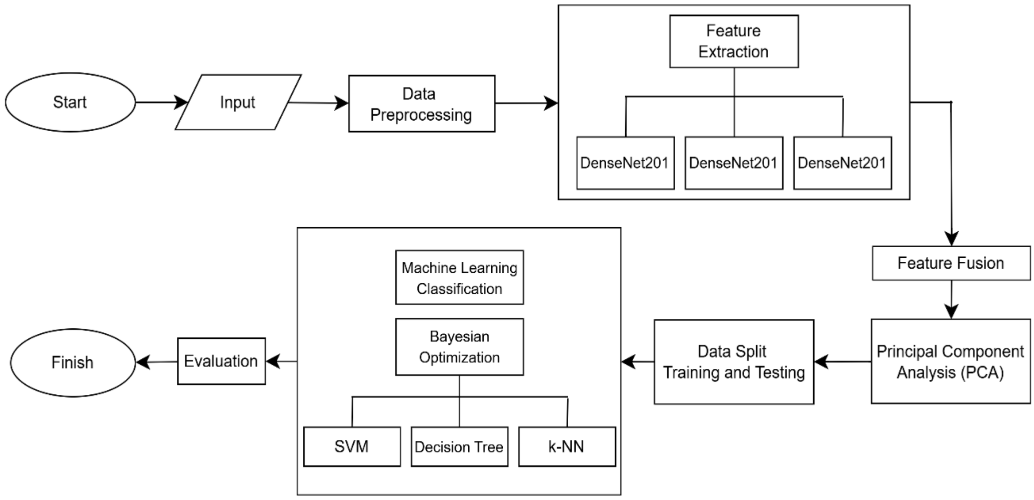

Figure 1. Research method

The resulting features are fused using one of the fusion rules, namely concatenate. The results of the feature fusion will be reduced in dimension using PCA. The data is then divided into training data and test data for classification using SVM, DT, and k-NN, with BO. In the evaluation phase of this study, performance indicators including accuracy, precision, recall, and F1-score are utilized to assess the impact of the implemented hybrid approach.

2.1. Dataset

Chest X-ray is a medical image that results from the process of radiography, which is an imaging technique that uses X-rays to visualize the internal structures of the chest. All images have been pre-processed and scaled to 224×224 pixels. The dataset was created by merging multiple datasets from the Kaggle platform, an open online repository for datasets. The datasets showed minimal imbalance but contained duplicate images, so dataset cleaning was performed. Table 1 presents the dataset both prior to and following the data cleaning procedure, while Figure 2 shows a sample chest X-ray image of the lungs. Subsequently, the dataset is randomly split into 85% for training and 15% for testing, ensuring that class distribution remains balanced. Training data serves to build the model, while test data is utilized to measure how well the model performs on unseen inputs.

Table 1. Number of images per class after data cleaning process

Classes	Normal	COVID-19	Pneumonia	Tuberculosis	Total
Before cleaning	1,802	1,626	1,800	1,600	6,828
After cleaning	1,671	1,537	1,791	1,600	6,599

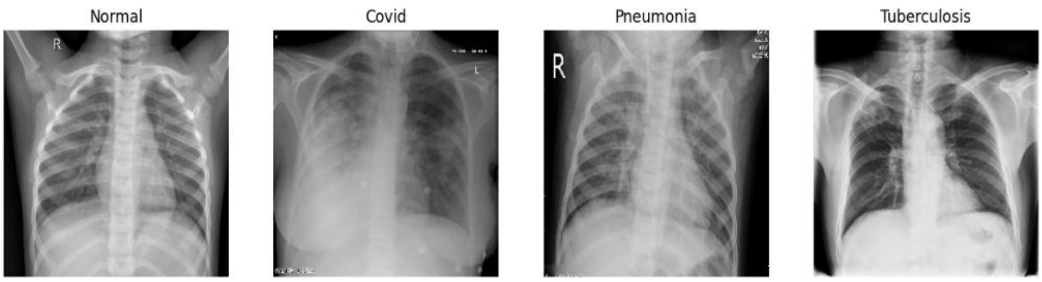

Figure 2. Chest X-ray of the lungs

2.2. Deep feature extraction

According to Heaton [21], CNN is a specific architecture in DL intended to process data organized in a grid format, such as picture data, represented as a 2D grid of pixels grouped in rows and columns. A CNN has layers of neurons that facilitate pattern recognition and feature extraction from pictures [22]. One of the main features of CNN is the convolution layer, this layer extracts characteristic visual information hierarchically from the input image. This procedure entails an input matrix that serves as a numerical depiction of the image alongside a filter, a diminutive matrix of a particular dimension. The filter functions as a feature detector, traversing the input matrix and executing an element-wise multiplication with the relevant segment of the input, with the products then summed to yield a single output value. The output values subsequently create a feature map that illustrates the characteristics identified by the filter. Stride is the step determining the magnitude of the filter shift; if the stride size is 1, the filter will advance by one pixel with each iteration. The convolution process may be mathematically expressed as (1), where $Y_{(i,j)}$ is the output value at position (i, j), A is the input matrix, K is the convolution filter of size $k \times k$, and s is the stride value.

$$Y_{(i,j)} = \sum_{m=1}^{k} \sum_{n=1}^{k} A_{(i-1)s+m,(j-1)s+n} \times K_{(m,n)} \quad (1)$$

Input images are processed by the pretrained model via a series of convolutional and pooling layers designed to extract hierarchical features. Each convolutional block augments the filter count, enabling the model to discern progressively intricate patterns. At the same time, the pooling layer reduces the dimensionality of the image. Max Pooling [23] selects the maximum activation value to represent the whole region, preserving essential information. This technique effectively generates a sparser representation by selecting only the highest activation value from each pooled region. This study employs three pre-trained CNN models: VGG16, InceptionV3, and DenseNet201, with an input picture size of 224×224. Figure 3 presents the structural design of each respective model.

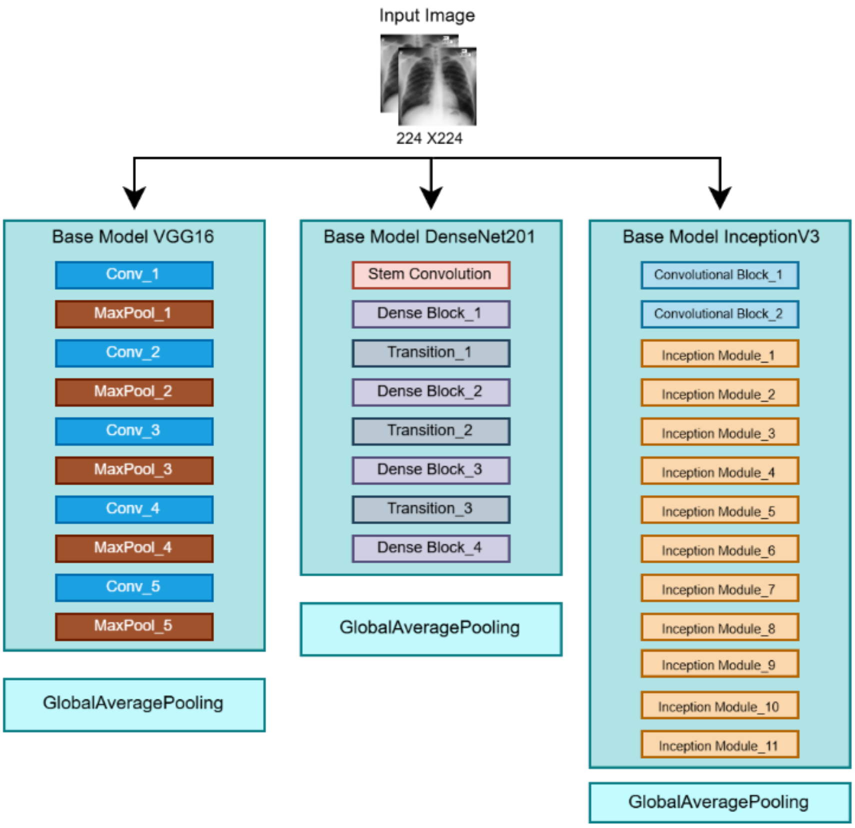

Figure 3. CNN model architecture

As shown in Figure 3, each model used integrates global average pooling as an additional layer after the main model. Global average pooling reduces data dimensions while retaining important information by utilizing the invariance of feature map averages, thereby reducing overfitting and improving computational efficiency. The VGG16 model has a stable and deep architecture with a straightforward sequential structure, making it suitable for extracting robust spatial features.

It produces 512-dimensional feature outputs. DenseNet201, designed to optimize information and gradient flow through dense connections between layers, generates richer feature representations with up to

1,920 output channels. Meanwhile, InceptionV3 incorporates multiple convolutional filter sizes within the same module, allowing it to capture multi-scale features effectively. This model produces final features with a dimension of 2,048 [24]–[26].

2.3. Feature fusion and dimensionality reduction

One feature type is frequently insufficient to accommodate an extensive range of data in classification issues. In order to provide a more reliable and accurate representation of the data, techniques that integrate characteristics from several extractors are thus required [27]. In order to integrate all of the features into a one-dimensional (1D) vector form, flattening is necessary because each model generates feature representations in a three-dimensional tensor format.

This guarantees that characteristics from several models may be consistently blended. The feature representations for each sample are flattened by transforming them from high dimensions h, w, c into 1D vectors with a single dimension of $h \cdot w \cdot c$. This procedure may be stated mathematically as (2). Besides in (2) and (3), i represents the model where n, h, w and c are the number of samples, feature height, width, and channels. After averaging, the features of the three models are fused using the concatenation method, resulting in a feature fusion matrix (FF) of dimension $(n, d_{InceptionV3} + d_{DenseNet201} + d_{VGG16})$.

$$F_{flat,i} \in \mathbb{R}^{n \times d_i} \quad (2)$$

Were,

$$d_i = h \cdot w \cdot c \quad (3)$$

Besides being widely used in modern network architectures for feature fusion, concatenation methods are used for their simplicity and ability to retain information from multiple feature sources without losing important details [28]. Unlike averaging or maximizing, which can reduce dimensionality or ignore variations between features, concatenation preserves the richness of the feature representation obtained from each model or layer. This fusion corresponds to the feature concatenation method expressed in (4).

$$FF = [F_{InceptionV3}, F_{VGG16}, F_{DenseNet201}] \quad (4)$$

However, the concatenation method increases the final feature vector's dimension, increasing computational complexity. Therefore, PCA is applied to the fused matrix (FF) to reduce the dimensionality of large and complex features while retaining the most important information, simplifying the representation of high-dimensional data. In general, PCA works by calculating the eigenvalues of the data covariance matrix C to determine the main direction of data dispersion through a decomposition process, as shown in (5).

$$C\boldsymbol{v} = \lambda \boldsymbol{v} \quad (5)$$

Pairs of eigen value λ_i and eigenvector v_i are obtained, which compose the principal components. These components are then sorted by the value of λ_i in descending order. This study chooses the minimum number of principal components k that satisfy (6).

$$\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{j=1}^{j} \lambda_j} \geq 0.95 \quad (6)$$

Where λ_i is the i-th eigenvalue representing the variance explained by the i-th principal component, and λ_j is the j-th eigenvalue of all d initial features, cumulatively representing the total variance of the original data. The 95% cumulative explained variance threshold is chosen to ensure that the reduced feature space retains most of the original information while significantly reducing dimensionality. By preserving at least 95% of the total variance, the transformed features remain highly representative of the original data structure, making them suitable for subsequent classification tasks.

2.4. Bayesian optimization for hyperparameter tuning in classification models

Classification was performed using three machine learning algorithms, SVM, DT, and k-NN, to explore different approaches and see the most effective model. BO was applied as a hyperparameter tuning method to improve classification performance. Hyperparameters [29] refer to settings that determine how a machine-learning model is trained. Hyperparameters are not learned from the data but are set before training and can significantly affect model performance. The basis of BO comes from Bayes' theorem in (7), i.e. when evaluating evidence data D, the posterior probability $P(H|D)$ of model H is equal to the probability $P(D|H)$ of evaluating evidence data D with model H multiplied by the prior probability $P(H)$.

$$P(H|D) \propto P(D|H)P(H) \tag{7}$$

The main objective of BO is to find the hyperparameter value x^* that maximizes the objective function $f(x)$, as shown in (8).

$$x^* = \arg\max_{x \in A} f(x) \tag{8}$$

In (8), x represents the combination of hyperparameters to be optimized. Meanwhile, $f(x)$ represents the objective function as the average accuracy score obtained from the training and evaluation process using k-fold cross-validation with a value of $k = 5$. BO in this study consists of two important elements. First, the Bayesian Gaussian processing (GP) statistical model is a probabilistic representation of the objective function. GP is a collection of random variables; each subset of random variables will follow a multivariate Gaussian distribution [30]. In (9) describes a Gaussian process through two core components: the mean function $m(x)$ and the covariance function $k(x, x')$.

$$f(x) \sim GP(m(x), k(x, x')) \tag{9}$$

The second element is the expected improvement (EI) acquisition function, which determines the next search point in hyperparameter space. EI is a method used to evaluate how much potential improvement there is at a point in optimization [31]. The main idea is to evaluate how much potential improvement in the objective function value can be compared to the current best value (f^*).

In its implementation, each machine learning algorithm's BO begins with a random selection of hyperparameter combinations as the starting point of exploration. Next, BO performs an optimization process for 100 iterations to explore the search space described in Table 2 and identify the hyperparameter combination that produces the best performance. The initial combination and the final result of the optimization process are comprehensively presented in Table 3. These optimized hyperparameters are utilized in both the training and testing processes to enhance classification model performance.

Table 2. Search space

ML alg.	Hyperparameter	Search space	Prior distribution
SVM	Box constraints (C)	[1e-5, 1e+1]	Log-uniform
	Kernel type	['linear', 'poly', 'rbf', 'sigmoid']	Categorical
	Kernel scale	[1e-3, 1e+1]	Log-uniform
	Polynomial degree	[2, 5]	Uniform (discrete)
	Kernel coefficient	[0, 1]	Uniform
DT	Maximum tree depth	[1, 50]	Log-uniform
	Minimum split size	{2, 7, 12, ..., 197}	Categorical
	Minimum leaf size	[1, 100]	Uniform
	Minimum leaf weight	[1e-2, 5e-1]	Log-uniform
	Splitting criterion	{‘gini’, ‘entropy’}	Categorical
K-NN	Number of neighbors	[1, 20]	Uniform
	Weights	{‘uniform’, ‘distance’}	Categorical
	Search algorithm	{‘auto’, ‘ball_tree’, ‘kd_tree’, ‘brute’}	Categorical

Table 3. Hyperparameter settings for machine learning models

ML alg.	Hyperparameter	VGG-16	DenseNet201	InceptionV3	Feature fusion
SVM	Box constraints (C)	4.97883	10	4.42014	10
	Kernel type	Linear	RBF	RBF	Poly
	Kernel scale	-	0.001	0.001	0.001
	Polynomial degree	-	-	-	2
	Kernel coefficient	-	-	-	1
DT	Maximum tree depth	50	50	37	23
	Minimum split size	7	7	2	87
	Minimum leaf size	1	1	1	26
	Minimum leaf weight	0.01	0.01	0.01104	0.01
	Splitting criterion	Entropy	Entropy	Gini	Entropy
K-NN	Number of neighbors	7	8	3	3
	Weights	Distance	Distance	Distance	Distance
	Search algorithm	Ball Tree	Kd Tree	Kd Tree	Ball Tree

3. RESULTS AND DISCUSSION

3.1. Feature enhancement result

Three different CNN-based models were employed to extract features from the processed lung images. The resulting outputs from each model were fused using concatenation to increase the robustness and diversity of the feature representation. Feature fusion yields high-dimensional data, which may lead to issues such as computational complexity. To mitigate these challenges, PCA is employed to project the features into a new subspace with reduced dimensions. The outcomes of this series of methodologies are displayed in Table 4.

Table 4. Feature extraction and dimensionality reduction results

Feature	VGG16	DenseNet201	InceptionV3	FF	FF-PCA
Feature dimension	512	1,920	2,048	4,480	1,217
Method	Global average pooling	Global average pooling	Global average pooling	Concatenation	Principal component analysis

The table indicates that each CNN model generates a distinct quantity of features. VGG16 generates 512 features, DenseNet201 generates 1,920 features, and InceptionV3 generates 2,048 features. InceptionV3 produces the most features, signifying that the model extracts more comprehensive information than VGG16 and DenseNet201. Following the concatenation of features, the dimensionality increased to 4,480. PCA substantially lowered the feature dimension from 4,480 to 1,217 features.

3.2. Evaluation of classification

Classification tasks were carried out using three machine learning methods, namely SVM, Decision DT, and k-NN, with the specific hyperparameters outlined in Table 3. The performance of each model was assessed using evaluation metrics such as accuracy, precision, recall, and F1 score, as represented in (10)-(13). The results derived from the evaluation are summarized in Table 5.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{10}$$

$$Precision = \frac{TP}{TP+FP} \tag{11}$$

$$Recall = \frac{TP}{TP+FN} \tag{12}$$

$$F1 - Score = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{13}$$

The model evaluation uses standard metrics such as true positive (TP), true negative (TN), false positive (FP), and false negative (FN) for performance assessment. Based on Table 5, the hybrid approach based on feature fusion (FF) with the combination of FF-SVM shows the best classification performance compared to single models such as VGG16, DenseNet201, and InceptionV3. FF-SVM recorded the highest accuracy of 98.79%, along with precision, recall, and F1-score, all above 98%. These results reflect the stability of the model in classifying multiclass medical images. Compared to baselines such as DenseNet201-SVM (97.47%) and VGG16-SVM (97.07%), feature fusion from multiple CNNs provides richer information.

Table 5. Evaluation result

Model	ML algoritms	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Time taken (s)
VGG16	SVM	97.07	97.16	97.17	97.16	244.6789
	DT	83.33	83.81	83.44	83.53	310.1874
	K-NN	95.86	95.91	96.00	95.94	202.0312
DenseNet201	SVM	97.47	97.55	97.46	97.49	648.1935
	DT	88.59	88.73	88.73	88.69	520.6981
	K-NN	96.46	96.59	96.48	96.51	727.9218
InceptionV3	SVM	95.45	95.54	95.53	95.53	1008.9276
	DT	79.80	79.88	79.83	79.81	678.3621
	K-NN	91.52	91.62	91.66	91.62	715.0327
FF	SVM	98.79	98.85	98.82	98.84	638.0538
	DT	88.79	88.85	88.88	88.84	508.6961
	K-NN	97.27	97.40	97.29	97.34	401.2894

Feature fusion also improves the performance of other algorithms. The accuracy of k-NN increased from 95.86% to 97.27%, and DT slightly increased from 88.59% to 88.79%. However, DT still showed the lowest performance among the three algorithms, indicating its limitations in recognizing complex visual patterns. Feature fusion resulting in high-dimensional feature vectors shows that the application of PCA significantly improves computational efficiency without degrading model accuracy. This effectiveness is reflected in the relatively efficient training time of the FF-SVM model, which is about 638 seconds. This duration is comparable to a single model such as DenseNet201-SVM, which takes 648 seconds of training time and is significantly faster than InceptionV3-SVM, which takes up to 1,008 seconds.

Pipeline efficiency is also improved by applying BO in the hyperparameter adjustment process. This approach offers advantages over conventional trial-and-error methods, as it is able to identify the optimal combination of parameters with fewer iterations automatically. As a result, the model training process becomes faster, and the computational burden can be reduced significantly. This is crucial, especially in system implementation in clinical environments that demand high efficiency and reliability.

According to the outcomes of the error analysis, the confusion matrix reveals that most instances were classified correctly, reflecting strong overall model performance. However, some misclassification cases remain, especially in classes with high visual similarity. In the proposed FF-SVM model, for example, there are 5 cases of tuberculosis, 4 cases of pneumonia classified as COVID-19, and one normal case classified as pneumonia. This finding reflects the challenges in distinguishing diseases that have overlapping radiological characteristics. Figure 4 presents a visual representation of the misclassification pattern through the confusion matrix for each model, which provides a comprehensive overview of the distribution of predictions between classes.

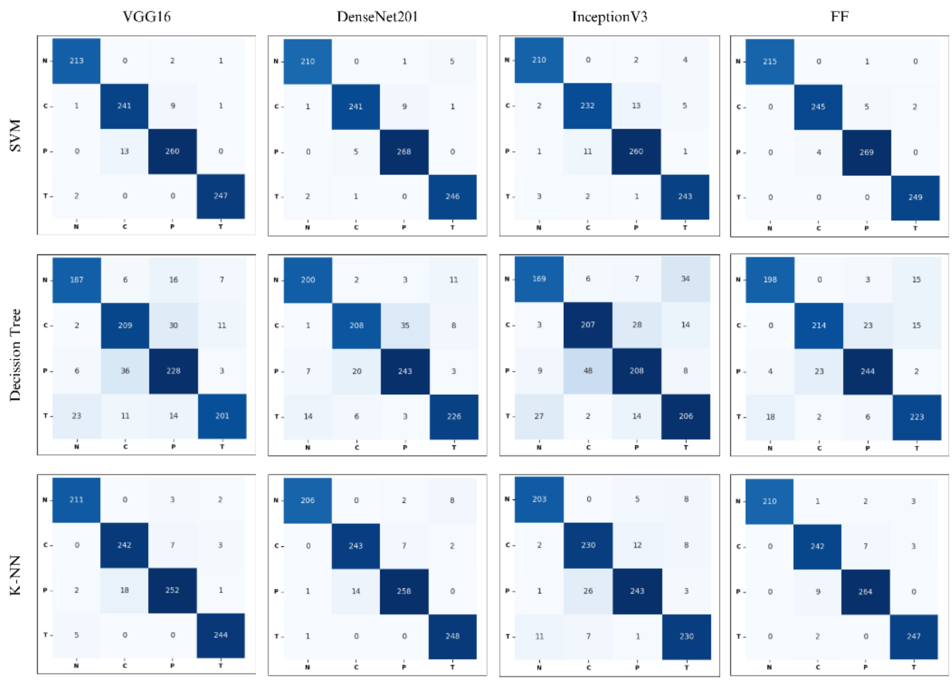

Figure 4. Confusion matrices for different classifiers and architectures

From the receiver operating characteristic (ROC) curves in Figure 5, FF-SVM recorded AUC =1.00 for all classes, indicating a very high classification ability. DT and k-NN were slightly lower, with average AUCs of 0.98 and 0.99, respectively. These results confirm the superiority of SVM, especially in distinguishing classes such as TB, as indicated by its perfect AUC of 1.00 across all classes.

The proposed method's superiority is also apparent compared to previous studies in Table 6. Previous approaches, such as VGGNet and ABO-CNN, only achieved 95.11% and 96.95% accuracy. Even modern ensemble approaches such as VGG-19 + Vision Transformer only yielded 94.52% accuracy. With an accuracy of 98.79%, the FF-SVM method surpasses these results and offers better time efficiency and performance stability. These results highlight the proposed approach's substantial contributions and relevance for chest X-ray image-based lung disease detection systems.

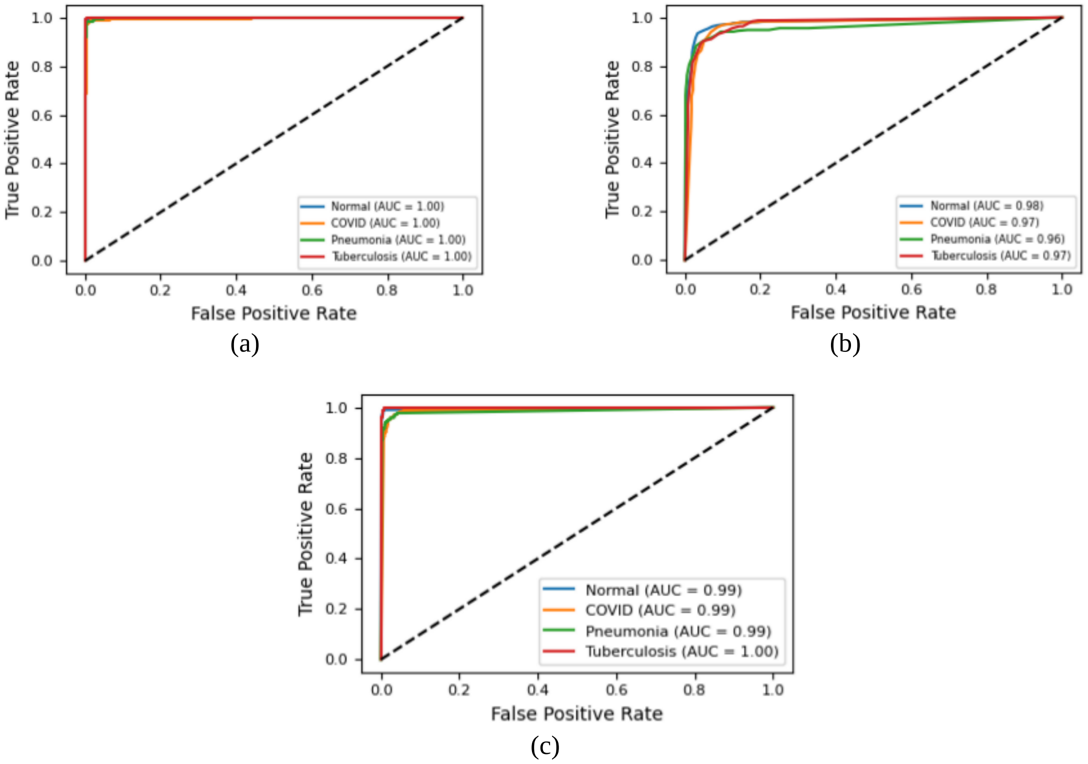

Figure 5. ROC curve of FF models (a) Support vector machine, (b) Decision tree, and (c) K-nearest neighbors

Table 6. Comparison with previous studi

Previous study	Methods	Dataset	Accuracy (%)
[32]	VGGnet	Chest X-Ray	95.11
[33]	Model Ensemble (VGG-19, ResNet50, vision transformer)	Chest X-Ray	92.46
			94.52
[34]	ABO-CNN	Chest X-Ray	96.95

4. CONCLUSION

This study shows that fusing features from several CNN-based architectures contributes to improved classification accuracy in CXR image analysis using machine learning models. The developed FF-SVM model achieved the highest accuracy of 98.79% in classifying four lung disease categories: Normal, COVID-19, Pneumonia, and Tuberculosis. These results indicate the great potential of this approach to be applied in medical diagnosis support systems, especially in applications that are easily accessible in areas with limited health facilities. However, this study has some limitations. The model was trained using only CXR images without considering other medical data such as CT scans, MRIs, or additional clinical data. In addition, potential biases in the dataset and reliance on image quality may limit the model's generalization to a broader population. Testing has not been conducted on external datasets nor in the context of real use in a clinical environment. As a development direction, exploring multimodal data fusion strategies and validation in real-world scenarios could be an important step to improve reliability and broader applicability. The methodology also supports the use of artificial intelligence (AI) systems in healthcare, which can be extended to other medical imaging modalities and different disease categories.

ACKNOWLEDGMENTS

During this research, the authors thank our friends for their encouragement, support, and motivation. Thanks to their tremendous assistance, the authors have been able to stay motivated and focused. The authors also acknowledge the invaluable advice, knowledge, and suggestions from our supervisors, which significantly raised the caliber of our study. Additionally, the authors are grateful to the journal editors and reviewers for their helpful criticism and recommendations, which enhanced the manuscript's consistency and clarity.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Dewi Rismawati	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓			
Sugiyarto Surono		✓		✓			✓			✓		✓	✓	✓
Aris Thobirin	✓				✓				✓	✓	✓	✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are openly available in kaggle at the following links:

– Dataset 1- COVID-19, Pneumonia, and Normal CXR Images
– Dataset 2- Tuberculosis CXR (Shenzhen)
– Dataset 3- Tuberculosis CXR (Montgomery)
– Dataset 4- TBX11 CXR (usmanshams)

REFERENCES

[1] World Health Organization, *World health statistics 2024: monitoring health for the SDGs sustainable development goals*. Geneva, Switzerland: World Health Organization, 2024.
[2] S. K. Zhou *et al.*, "A review of deep learning in medical imaging: imaging traits, technology trends, case studies with progress highlights, and future promises," in *Proceedings of the IEEE*, 2021, vol. 109, no. 5, pp. 820–838.
[3] W. Ma and B. Guo, "Advanced neurological activity status of athletes based on big data technology," *Heliyon*, vol. 10, no. 17, 2024.
[4] M. F. Ijaz and M. Woźniak, "Editorial: Recent advances in deep learning and medical imaging for cancer treatment," *Cancers*, vol. 16, no. 4, 2024.
[5] S. T. H. Kieu, A. Bade, M. H. A. Hijazi, and H. Kolivand, "A survey of deep learning for lung disease detection on medical images: state-of-the-art, taxonomy, issues and future directions," *Journal of Imaging*, vol. 6, no. 12, 2020.
[6] G. Gautam and A. Khanna, "Content based image retrieval system using CNN based deep learning models," in *Procedia Computer Science*, vol. 235, pp. 3131–3141, 2024.
[7] G. Hamed, M. A. E. R. Marey, S. E. S. Amin, and M. F. Tolba, "Deep learning in breast cancer detection and classification," *Advances in Intelligent Systems and Computing*, pp. 322–333, 2020.
[8] A. B. Ramakrishnan, M. Sridevi, S. K. Vasudevan, R. Manikandan, and A. H. Gandomi, "Optimizing brain tumor classification with hybrid CNN architecture: balancing accuracy and efficiency through oneAPI optimization," *Informatics in Medicine Unlocked*, vol. 44, 2024.
[9] K. Gupta and N. Chawla, "Analysis of histopathological images for prediction of breast cancer using traditional classifiers with pre-trained CNN," in *Procedia Computer Science*, 2020, vol. 167, pp. 878–889.
[10] M. F. Aslan, K. Sabanci, A. Durdu, and M. F. Unlersen, "COVID-19 diagnosis using state-of-the-art CNN architecture features and Bayesian optimization," *Computers in Biology and Medicine*, vol. 142, 2022.
[11] A. Biswas and M. S. Islam, "A hybrid deep CNN-SVM approach for brain tumor classification," *Journal of Information Systems Engineering and Business Intelligence*, vol. 9, no. 1, 2023.
[12] X. Zhao, C. Tang, H. Hu, W. Wang, S. Qiao, and A. Tong, "Attention mechanism based multimodal feature fusion network for human action recognition," *Journal of Visual Communication and Image Representation*, vol. 110, 2025.
[13] I. T. Jolliffe, "Principal component analysis," in *Springer series in statistics*, 2nd ed., New York, USA: Springer, 2002.
[14] S. A. Salleh *et al.*, "Support vector machine (SVM) and object-based classification in earth linear features extraction: a comparison," *Revue Internationale de Géomatique*, vol. 33, no. 1, pp. 183–199, 2024.
[15] J. W. Han, M. Kamber, and J. Pei, "Classification: basic concepts," in *Data mining: concepts and techniques*, Burlington, Massachusetts: Morgan Kaufmann, 2012.
[16] Z. E. Rasjid and R. Setiawan, "Performance comparison and optimization of text document classification using k-NN and naïve Bayes classification techniques," *Procedia Computer Science*, vol. 116, pp. 107–112, 2017.
[17] Y. Yao, J. Cao, and Z. Ma, "A cost-effective deadline-constrained scheduling strategy for a hyperparameter optimization workflow for machine learning algorithms," in *Service-Oriented Computing*, vol. 11236, New York, USA: Springer, Cham, 2018.

[18] Q. A. Hidayaturrohman and E. Hanada, "A comparative analysis of hyper-parameter optimization methods for predicting heart failure outcomes," *Applied Sciences*, vol. 15, no. 6, 2025.
[19] A. Alzahem, W. Boulila, M. Driss, and A. Koubaa, "Feature fusion for improved classification: combining dempster-shafer theory and multiple CNN architectures," in *Computational Collective Intelligence*, vol. 14811, New York, USA: Spriger cham, 2024.
[20] M. Zhang, H. Li, J. Lyu, S. H. Ling, and S. Su, "Multi-level CNN for lung nodule classification with Gaussian process assisted hyperparameter optimization," *arXiv Computer Science*, 2019.
[21] J. Heaton, "Ian Goodfellow, Yoshua Bengio, and Aaron Courville: deep learning," *Genetic Programming and Evolvable Machines*, vol. 19, no. 1–2, pp. 305–307, 2018.
[22] M. M. Mashamba, A. Telukdarie, I. Munien, U. Onkonkwo, and A. Vermeulen, "Detection of bacterial spot disease on tomato leaves using a convolutional neural network (CNN)," in *Procedia Computer Science*, vol. 237, pp. 602–609, 2024.
[23] Y. Doğan, "Which pooling method is better: max, avg, or concat (max, avg)," *Communications Faculty of Sciences University of Ankara Series A2-A3 Physical Sciences and Engineering*, vol. 66, no. 1, pp. 95–117, 2024.
[24] N. M. Blauch, M. Behrmann, and D. C. Plaut, "Computational insights into human perceptual expertise for familiar and unfamiliar face recognition," *Cognition*, vol. 208, 2021.
[25] M. K. A. Mazumder, M. M. Kabir, A. Rahman, M. Abdullah-Al-Jubair, and M. F. Mridha, "DenseNet201Plus: Cost-effective transfer-learning architecture for rapid leaf disease identification with attention mechanisms," *Heliyon*, vol. 10, no. 15, 2024.
[26] M. J. Ferdous and R. Shahriyar, "An ensemble convolutional neural network model for brain stroke prediction using brain computed tomography images," *Healthcare Analytics*, vol. 6, 2024.
[27] T. Akilan, Q. J. Wu, and H. Zhang, "Effect of fusing features from multiple DCNN architectures in image classification," *IET Image Processing*, vol. 12, no. 7, pp. 1102–1110, 2018.
[28] Y. Dai, F. Gieseke, S. Oehmcke, Y. Wu, and K. Barnard, "Attentional feature fusion," *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 3559–3568, 2021.
[29] L. T. Aleksina, "Hyperparameters optimization for the machine learning," *Connectivity*, vol. 168, no. 2, 2024.
[30] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning*. Massachusetts, USA: MIT Press, 2006.
[31] J. Wu, X. Y. Chen, H. Zhang, L. D. Xiong, H. Lei, and S. H. Deng, "Hyperparameter optimization for machine learning models based on Bayesian optimization," *Journal of Electronic Science and Technology*, vol. 17, no. 1, pp. 26–40, 2019.
[32] S. Park *et al.*, "Enhancing chest X-ray image classification for lung diseases through machine learning," *Journal of Student Research*, vol. 12, no. 3, 2023.
[33] M. Pal, S. Parija, and G. Panda, "An effective ensemble approach for classification of chest X-ray images having symptoms of COVID: a precautionary measure for the COVID-19 subvariants," *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, vol. 8, 2024.
[34] N. Alalwan, A. I. Taloba, A. Abozeid, A. I. Alzahrani, and A. H. Al-Bayatti, "A hybrid classification and identification of pneumonia using African Buffalo optimization and CNN from chest X-ray images," *CMES - Computer Modeling in Engineering and Sciences*, vol. 138, no. 3, pp. 2497–2517, 2024.

BIOGRAPHIES OF AUTHORS

Dewi Rismawati is an undergraduate student in the Mathematics Study Program at the Faculty of Applied Science and Technology, Universitas Ahmad Dahlan. Through her academic background, she has developed a strong interest in machine learning and deep learning. She is actively involved in academic and research activities focusing on the application of artificial intelligence technologies to data-driven problem solving. She can be contacted at email: dewi2100015008@webmail.uad.ac.id.

Sugiyarto Surono is a Professor at the Mathematics Study Program, Faculty of Applied Science and Technology, Ahmad Dahlan University. His research interests include big data, machine learning, deep learning, and artificial intelligence. Throughout his academic career, he has been actively involved in research and development in the fields of data science and computational intelligence. His contributions focus on the development of mathematical models and intelligent systems, aiming to integrate mathematical theory with real-world applications. He is dedicated to advancing mathematics education and research through modern computational approaches. He can be contacted at email: sugiyarto@math.uad.ac.id.

Aris Thobirin is a senior lecturer in the Mathematics Study Program, Faculty of Applied Science and Technology, Universitas Ahmad Dahlan. His areas of interest include analysis, algebra, and machine learning. With a strong background in both pure and applied mathematics, he is actively engaged in teaching and research, particularly in bridging mathematical theory with its applications in modern computational technologies. He can be contacted at email: aris.thobi@math.uad.ac.id.