Bulletin of Electrical Engineering and Informatics
Vol. 15, No. 2, April 2026, pp. 1417~1428
ISSN: 2302-9285, DOI: 10.11591/eei.v15i2.10585

Detecting respiratory diseases using spectrogram-based deep features and machine learning algorithms

Elvina Nur Hana[1], Mohammad Reza Faisal[1], Dwi Kartini[1], Muhammad Itqan Mazdadi[1], Setyo Wahyu Saputro[1], Fatma Indriani[1], Kenji Satou[2]
[1]Department of Computer Science, Faculty of Mathematics and Natural Sciences, Lambung Mangkurat University, Banjarbaru, Indonesia
[2]Faculty of Transdisciplinary Sciences for Innovation, Institute of Transdisciplinary Sciences for Innovation, Kanazawa University, Kanazawa, Japan

Article Info

Article history:

Received Apr 28, 2025
Revised Feb 24, 2026
Accepted Mar 5, 2026

Keywords:

Deep feature
Lung sound
Machine learning
Respiratory diseases
Spectrogram

ABSTRACT

Early diagnosis of respiratory diseases is difficult as lung sound analysis requires the skills of medical professionals. Respiratory diseases are one of the leading causes of death in the world, so early detection is critical. Automatic identification is made possible by artificial intelligence. However, lung sound data is unstructured, while artificial intelligence often requires structured data. Therefore, feature extraction is required to structure the voice data. Traditional techniques such as mel-frequency cepstral coefficients (MFCC) often produce fewer features and information. This research uses a deep feature approach, which produces more features, as a solution. This research applies three convolutional neural network (CNN) architectures as deep features, namely VGG-16, DenseNet-121, and ResNet-50, with machine learning classifications, namely random forest, support vector machine (SVM), Naïve Bayes, and K-nearest neighbors (KNN). This research will identify the optimal combination of methods. The results of this study show that respiratory disease classification can be effectively achieved by combining deep features and machine learning classification. The results of 10-fold cross-validation show that the three CNN architectures perform best on SVM with a linear kernel. The accuracy of VGG-16 is 70.63%, ResNet-50 is 64.93%, and DenseNet-121 is 73.58%.

Corresponding Author:

Mohammad Reza Faisal
Department of Computer Science, Faculty of Mathematics and Natural Sciences
Lambung Mangkurat University
Banjarbaru, Indonesia
Email: reza.faisal@ulm.ac.id

1. INTRODUCTION

According to the World Health Organization (WHO), respiratory diseases are one of the leading causes of death worldwide [1]. Therefore, early detection of respiratory diseases is essential to stop complications and speed up treatment. Stethoscopes are still used to identify respiratory conditions by listening for abnormal lung sounds in the anterior and posterior thoracic regions, such as crackles and wheezes [2]. These lung sounds provide critical information for the diagnosis of respiratory diseases. However, this conventional method is unsuitable for long-term monitoring, and manual auscultation relies heavily on the skill and hearing of the doctor, which can result in misdiagnosis [3]. With the development of technology, artificial intelligence-based early detection with lung sound input began to be developed to overcome these limitations, for example, in the research of Demir *et al.* [4] using convolutional neural

network (CNN) on spectrogram images of lung sounds generated through short-time Fourier transform (STFT). Also, Yang *et al.* [5] used lung sound input for respiratory disease classification and converted it into STFT and wavelet transform representations combined with BInet. Another research by Ali *et al.* [6] used a 1D-CNN approach to raw lung sound signals.

The development of artificial intelligence for respiratory disease classification using lung sound input shares similarities with other audio classification tasks. Generally, it consists of two main stages, i.e., feature extraction and classification. Feature extraction converts raw audio signals into structured data so that classification algorithms recognize relevant patterns [7]. Mel-frequency cepstral coefficients (MFCC) is a commonly used feature extraction technique, which efficiently reduces dimensionality but can lose some information in complex data that can affect accuracy [8]. On the other hand, a deep feature technique utilizes all features and produces deeper features, one of which is the CNN approach. For example, research [9] using the ResNet-50 deep feature method and a linear support vector machine (SVM) classifier achieved classification of skin disease types on a balanced image dataset, with an accuracy of 97% being reached. In another research [10], multilabel chest diseases were detected using X-ray images with the transfer learning method on DenseNet121, which served as a deep feature, and achieved an accuracy of up to 98.5% for diseases such as edema.

Additionally, in study [11], deep features were extracted from VGG-16 and classified using an ensemble of several machine learning algorithms to detect brain tumors, achieving an accuracy of up to 98.15%. This demonstrates that CNN methods effectively extract deep features in image classification. This study's data consists of audio signals that will be converted into spectrogram images. This approach is expected to provide similar performance when applied to audio data. Therefore, based on the good results reported in the aforementioned image classification studies, this study will use the ResNet-50 and VGG-16 techniques for deep feature extraction.

Once the structured data has undergone the feature extraction process, the next step is classification. Currently, machine learning and deep learning algorithms are commonly used. For example, the research of Rayan and Alaerjan [12] applied Bi-LSTM to classify COVID-19 from X-ray images with an accuracy rate of 93%. In another study, random forest achieved 92% accuracy, outperforming other algorithms in classifying chronic obstructive pulmonary disease (COPD) [13]. Meanwhile, Demir *et al.* [4] used a transfer learning approach with CNN and a SoftMax classifier, resulting in 65.5% and 63.09% accuracy in lung disease classification. Moreover, in Parkinson's disease diagnosis, MFCC with SVM kernel radial basis function (RBF) achieved 77.5% accuracy [14]. Although deep learning often shows high accuracy, the choice of algorithm depends on the complexity of the data and the need for interpretability because, in some cases, machine learning is more explainable than black-box deep learning [15]. In this study, classification is performed using several popular machine learning algorithms, considering the diversity of classification approaches. Random forest represents a decision tree-based ensemble approach [16]. Naïve Bayes is a probabilistic model that works well on data with independent feature assumptions [17]. SVM is a margin-based method that can form non-linear decision boundaries using kernels [18]. Meanwhile, K-nearest neighbors (KNN) is an instance-based method that classifies data based on the dimensions of the feature space [19].

This study took several key steps to address challenges commonly encountered in real-world data, including variations in signal data duration and data imbalance. Since the data obtained had different frequencies, the frequencies were resampled. Moreover, segmentation was also performed with varying data durations to facilitate further analysis. To address class imbalance, this study applied a data balancing method. Additionally, validation was performed using 10-fold cross-validation to ensure robust model evaluation and reduce the risk of overfitting [16]. Random undersampling is a simple and straightforward data balancing method that reduces the number of samples in the majority class to enhance feature space class separation and mitigate classification bias [17]. This study differs from previous studies regarding preprocessing, data balancing, and the methods applied. The main question addressed in this study is to what extent the combination of deep feature extraction using CNN architectures (VGG-16, ResNet-50, and DenseNet-121) and machine learning algorithms can achieve the best accuracy in classifying respiratory diseases based on lung sounds, as well as which classification algorithm provides the most optimal performance for each CNN architecture. This study aims to classify respiratory diseases based on lung sounds using a combination of CNN deep features and machine learning algorithms, and to comprehensively evaluate their accuracy.

The novelty of this study is its thorough and clear approach. Unlike earlier work that used only end-to-end CNN models or handcrafted features like MFCC, this study combines machine learning methods, such as SVM, random forest, Naïve Bayes, and KNN, with features from pre-trained CNNs, including VGG-16, ResNet-50, and DenseNet-121. By also incorporating segmentation, random undersampling, and testing with 10-fold cross-validation, this method effectively addresses real-world issues in the ICBHI 2017 dataset, such

as class imbalance and variations in signal frequency or duration. Beyond offering a comparative analysis of various CNN combinations with machine learning, this research method provides a reliable and reproducible method for classifying respiratory sounds in practical environments.

2. METHOD

The research stages used in this study are depicted in Figure 1.

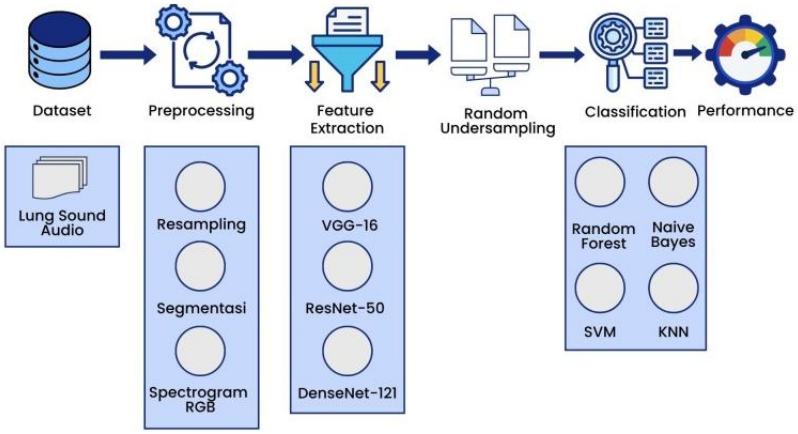


Figure 1. Diagram of the research workflow

2.1. Dataset

This study utilizes the ICBHI 2017 dataset on the Kaggle website. This dataset comprises audio recordings of respiratory sounds collected by researchers from Portugal and Greece. Annotations are made manually by respiratory specialists by marking the beginning and end of each breathing cycle and noting the presence of adventitious sounds: crackles and wheezes. There were 920 annotated sound recordings, with durations ranging from 10 to 90 seconds, collected from 126 patients. The audio duration is 5.5 hours and covers 6,898 breathing cycles taken from patients of different age groups. It also has four classes with unbalanced data. The distribution of data for each class is presented in Table 1.

Table 1. Distribution of datasets in each class

Class data	Amount
Normal	3,642
Crackles	1,864
Wheezes	886
Crackles and wheezes	506

2.2. Data preprocessing

Raw audio data is resampled to 4,000 Hz. Since abnormal lung sounds, such as crackles and wheezes, are typically below 2,000 Hz, resampling to 4,000 Hz can be performed. This is followed by fixed-duration segmentation of 2.7 seconds, corresponding to the average duration of a single breathing cycle. Segmentation is performed by cutting the beginning portion and adding sample padding if the duration is less than 2.7 seconds. This process aims to standardize the input data size, enabling the machine learning model to capture patterns more effectively [20].

After segmentation, the audio signal is converted into a visual representation using the STFT, which represents the frequency content of the audio segment window. The sound wave can be described as a two-dimensional function of time and frequency by building this representation over time. The square of the magnitude of the STFT representation, $|F(n,\omega)|^2$ is known as the spectrogram [4]. The spectrogram generated from the STFT is a visual representation of the signal's frequency content over time [21]. Figure 2 is the preprocessed dataset with a red, green, and blue (RGB) spectrogram [22]. The dataset is divided into four categories: the normal category illustrated in Figure 2(a), the crackles category depicted in Figure 2(b), the wheezes category represented in Figure 2(c), and the category encompassing both crackles and wheezes shown in Figure 2(d).

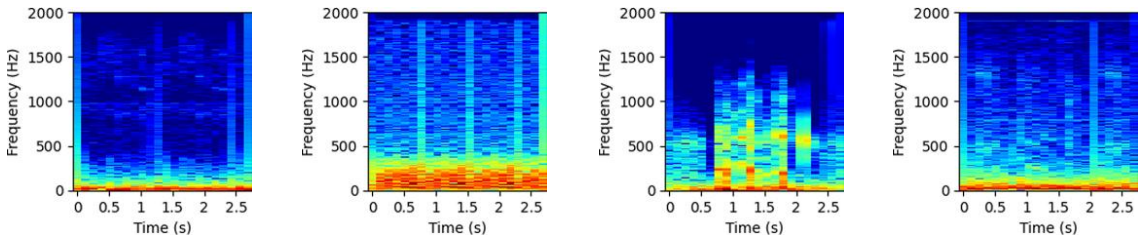


Figure 2. Example spectrogram representations of respiratory sound classes in the dataset; (a) normal class, (b) crackles class, (c) wheezes class, and (d) crackles and wheezes classes

2.3. Deep feature extraction

This study extracted deep features from spectrogram images generated through audio conversion using a CNN. The ImageNet dataset was used to train three CNN architectures—VGG-16, ResNet-50, and DenseNet-121—with an input resolution of 224×224 pixels. These three architectures were selected due to their distinct network structures. CNNs generally use two-dimensional (2D) convolution filters to extract high-level information from image inputs, making them suitable for analyzing RGB spectrogram images [23]. In the final stage, features from the final convolution layer are extracted and flattened into a one-dimensional vector, which is then used as input to the machine learning classification algorithm [24]. The characteristics of the CNN architecture are as follows:

- VGG-16: consists of 16 layers, including thirteen convolutional layers and three fully connected layers. This model employs relatively simple convolutional filters (3×3) and a deep network, making it easy to use and effective for image classification tasks [25]. The structure of VGG-16 is shown in Figure 3(a).
- DenseNet-121: each layer is designed using dense connectivity, where the layer receives input from all previous layers and provides output to all subsequent layers. This enhances the effectiveness of the gradient between layers and information propagation [26]. Figure 3(b) shows a graphical representation of this design.
- ResNet-50 has an architecture with two main layers: conv blocks and identity blocks. These two blocks enable gradients to flow more efficiently during training by forming shortcut connections between layers, as illustrated in Figure 3(c) [27].

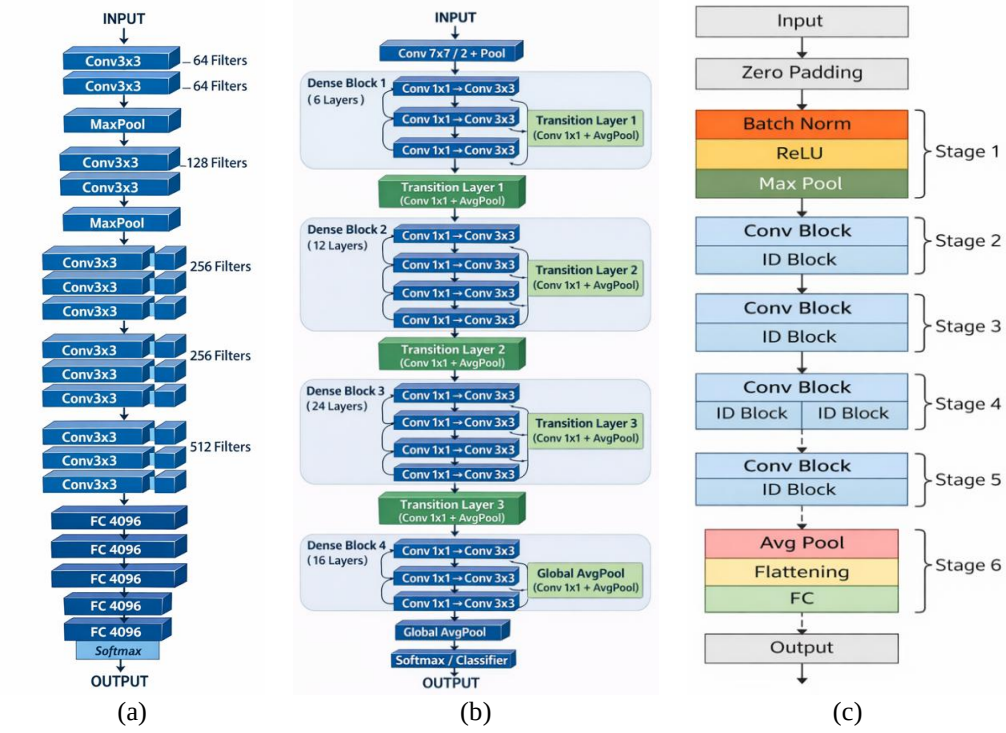


Figure 3. Visualization of the CNN architectures used for deep feature extraction; (a) architecture of VGG-16 [28], (b) architecture of DenseNet-121 [29], and (c) ResNet identity block and conv block structure [27]

2.4. Random undersampling

The method addresses data imbalance by randomly removing instances from the majority class until the number of cases in both the majority and minority classes is equal [30]. In this research, we applied random undersampling with stratified sampling to generate balanced subsets containing 500 samples per class.

2.5. 10-fold cross-validation

In this study, using 10-fold cross-validation, the 10-fold technique divides the network input data into ten subgroups. For each iteration of model training, nine subgroups are used as training data, while the remaining subgroups are used as test data [31], [32]. The formula for calculating the performance metric across all folds is as (1):

$$CV_{10} = \frac{1}{10} \sum_{i=1}^{10} MSE \quad (1)$$

2.6. Classification

After feature extraction, the machine learning model is classified using the Scikit-Learn library, which implements all classification functions with default parameters.

- Random forest: the decision tree algorithm is the evaluator of this algorithm. The principle is based on the bootstrap aggregating (Bagging) algorithm and random feature selection [33].
- SVM using various kernel functions to classify non-linear data sets, some SVM kernels are linear, polynomial, sigmoid, and RBF [34]. The functions of the SVM kernels used are presented in Table 2.

Table 2. Function mathematics kernel SVM

Kernel	Function
Linear	$k(x_i.x) = x_i.x$
Polynomial	$k(x_i.x) = (\gamma.x_i + C)^d$
Sigmoid	$k(x_i.x) = \tan h\ (\gamma.x_i + C)$
RBF	$k(x_i.x) = \exp(-\gamma\|x_i - x\|^2)$

- Naïve Bayes: serves to identify or predict the probability of future events (posterior probability). i.e., the possibility of an unknown data class based on information about previous events (prior probability, possibility, and evidence) [35].
- KNN: performs class classification based on the nearest neighbor distance. The Euclidean distance is often used to determine the distance between two objects, x and y [36].

2.7. Evaluation matrix

In this study, the evaluation metric used is the confusion matrix, which will be employed to assess the effectiveness of the machine learning classification model. The confusion matrix generated by each classification model is used to calculate the parameters for accuracy and F1-score [37], [38]. Since this research addresses a multiclass classification problem, AUC and ROC metrics for binary classification are not applied in this study [39]. The equation for performance in this multiclass study is as (2) and (3) [40]:

$$\text{Accuracy} = \frac{Number\ of\ correct\ predictions}{Total\ sample} \quad (2)$$

$$\text{F1-score macro} = \frac{1}{N} \sum_{i=1}^{N} F1_i \quad (3)$$

2.8. Experimental setup

The experimental configuration consisted of deep feature extraction using pre-trained CNN architectures and machine learning classification algorithms. Table 3 details the architectural configurations for feature extraction, while Table 4 summarizes the hyperparameter settings for the classification models. All models were trained on balanced subsets generated through random undersampling.

Table 3. Experiment setup on feature extraction

Parameter	Parameter value		
	VGG-16	ResNet-50	DenseNet-121
architecture	VGG-16	ResNet-50	DenseNet-121
weights	imagenet	imagenet	imagenet
Input_shape	224, 224, 3	224, 224, 3	224, 224, 3
imageDataGenerator	rescale=1./255	rescale=1./255	rescale=1./255
batch_size	128	128	128
output layer	block5_pool	conv5_block3_out	relu

Table 4. Experiment setup for classification

Machine learning	Algorithm	Parameter	Value
Random forest	RandomForestClassifier	n_estimators	100
		random_state	42
Naïve Bayes	GaussianNB	-	-
SVM linear	SVC	kernel	linear
		random_state	42
SVM polynomial	SVC	kernel	poly
		degree	3
		random_state	42
SVM RBF	SVC	kernel	rbf
		random_state	42
SVM sigmoid	SVC	kernel	sigmoid
		random_state	42
KNN (k=1)	KNeighborsClassifier	n_neighbors	1
KNN (k=3)	KNeighborsClassifier	n_neighbors	3
KNN (k=5)	KNeighborsClassifier	n_neighbors	5

3. RESULTS AND DISCUSSION

3.1. Results

Three CNN architectures have resulted in feature vectors with varying dimensions, as shown in Table 5. Table 6 shows the performance of each combination of CNN architecture and machine learning algorithm.

Table 5. The number of deep feature results

Architecture CNN	Feature count
VGG-16	25.088
ResNet-50	100.352
DenseNet-121	50.176

Table 6. Metric evaluation results

Machine learning	Deep feature					
	VGG-16		ResNet-50		DenseNet-121	
	Acc (%)	F1-score (%)	Acc (%)	F1-score (%)	Acc (%)	F1-score (%)
Random forest	68.08	67.57	62.17	61.47	64.93	64.45
Naïve Bayes	48.67	47.95	43.32	39.74	45.72	44.87
SVM (linear)	**70.63**	70.51	**64.93**	64.73	**73.58**	73.37
SVM (poly)	53.57	53.32	56.17	54.98	60.53	59.50
SVM (RBF)	66.37	65.63	59.52	58.58	70.23	69.86
SVM (sigmoid)	67.23	66.67	51.37	50.03	68.48	67.96
KNN=1	62.29	62.02	58.27	58.24	68.88	68.73
KNN=3	57.67	57.54	56.22	56.02	67.23	66.97
KNN=5	58.12	57.69	54.97	54.54	66.38	65.89

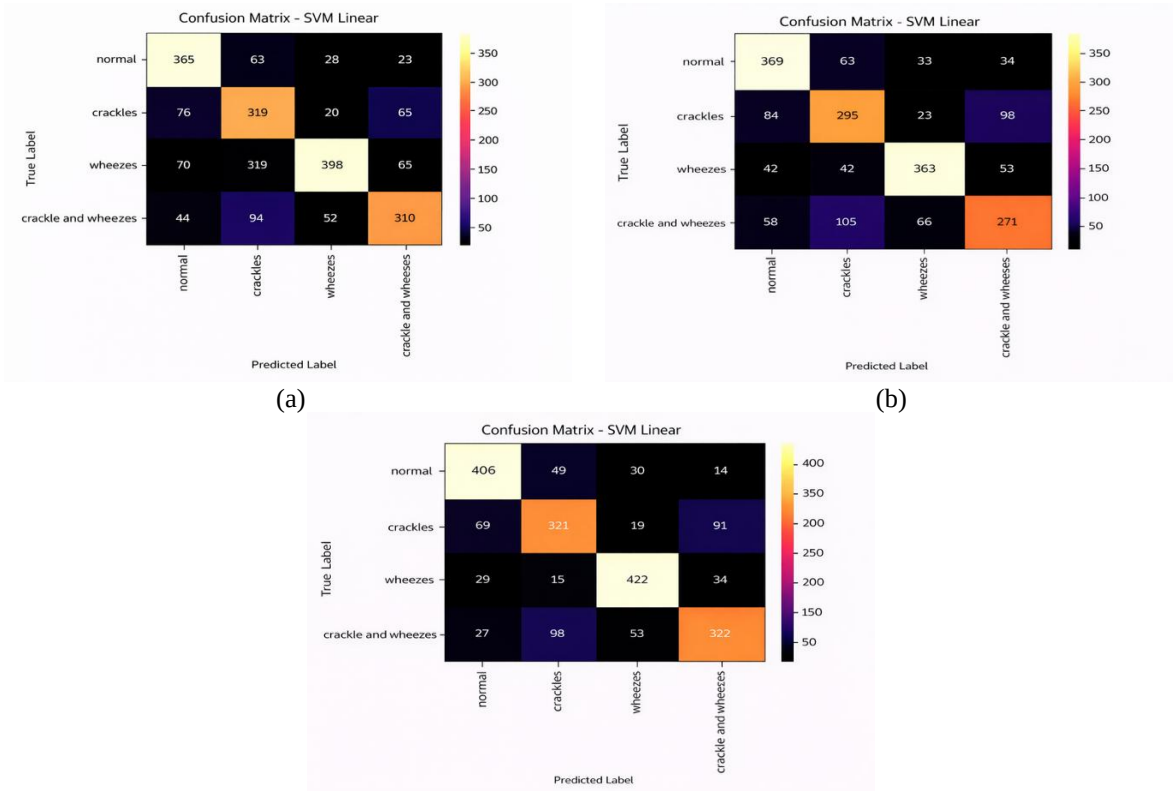
Based on the table above, the VGG-16, ResNet-50, and DenseNet-121 architectures produce the best combination in the linear kernel SVM method. Figure 4 shows the confusion matrix for each combination of the most optimized deep feature and machine learning. Figure 4(a) represents the confusion matrix outcome of a SVM model that utilizes inputs derived from the VGG-16 feature. Figure 4(b) illustrates a confusion matrix outcome from an SVM model that employs inputs sourced from the ResNet-50 feature. Figure 4(c) delineates the confusion matrix outcome of the SVM model utilizing inputs extracted from the DenseNet-121 feature. Based on the confusion matrix, the three CNN architectures with linear kernel SVM have superior predictive performance in the normal and wheezes classes, yielding more accurate predictions than the other classes.

3.2. Discussion

For the overall performance analysis, the average accuracy of each CNN architecture is calculated and displayed as a chart in Figure 5(a). DenseNet-121 produces the best performance, with an average accuracy of 65.11%, followed by VGG-16, and the lowest is ResNet-50. This difference could be due to how well each architecture extracts relevant features. ResNet-50 tends to overgenerate irrelevant features, thus reducing accuracy. On the other hand, VGG-16 has a few features that do not provide much information.

DenseNet-121 shows a better balance in feature extraction, but accuracy is still not good enough; with the combination of a linear kernel, SVM can improve classification performance.

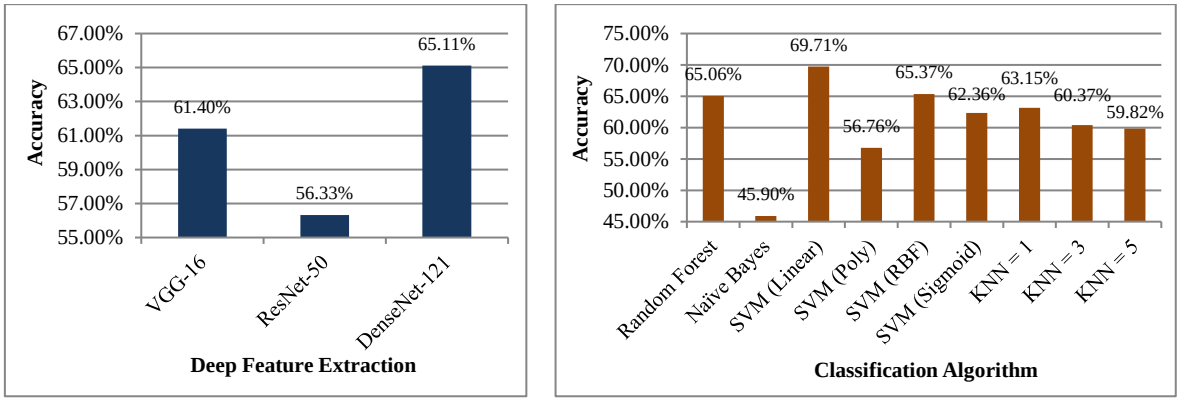
Meanwhile, the average performance of each machine learning algorithm shown in Figure 5(b) indicates that SVM with a linear kernel has the best performance with an average result of 69.71% compared to the other methods, which shows the ability of SVM to utilize the data generated by CNN.



(a) (b)

(c)

Figure 4. Confusion matrix showing the categorization outcomes of the top-performing CNN architecture combinations; (a) confusion matrix SVM kernel liner on VGG-16, (b) confusion matrix SVM linear on ResNet-50, and (c) confusion matrix SVM kernel linear on DenseNet-121



(a) (b)

Figure 5. Comparison of average classification performance across different model configurations; (a) average accuracy based on deep feature models and (b) average accuracy based on machine learning classification models

In this study, the accuracy of the results obtained in each classification method is still below 80%. This study is due to the many features generated during deep feature extraction, which can cause noise and unimportant information. Reducing the performance of machine learning models [41]. Furthermore, because the model must process and interpret high dimensions. An overabundance of features also results in a very long computation time.

To assess whether the observed performance differences were statistically significant, an unpaired t-test was conducted on the F1-scores obtained from both the feature extraction stage and the classification stage. The unpaired test was chosen because the compared groups represent independent cross-validation results. For CNN-based feature extraction, the comparison showed that DenseNet-121 did not differ significantly from VGG-16 ($p=0.3432$), but when compared to ResNet-50, it was statistically significant ($p=0.0229$). Meanwhile, for machine learning classifiers, the results indicated that the difference between SVM with a linear kernel and SVM with an RBF kernel was not significant ($p=0.3084$), and the difference between SVM linear and random forest was also not significant ($p=0.1784$). These findings suggest that DenseNet-121 provides a meaningful advantage over ResNet-50 in feature extraction, while among the top-performing classifiers, the superiority of SVM linear over SVM RBF and random forest cannot be considered statistically conclusive.

Table 7 presents a comparison of the performance of this approach with previous studies. The study by Demir *et al.* [4] employed a relatively simple transfer learning approach using CNNs with a SoftMax classifier, which achieved accuracies of 63.09% and 65.5%. However, the small size of the ICBHI 2017 dataset and the noisy, heterogeneous recordings' generalization ability of the end-to-end CNN model. On the other hand, the study by Yang *et al.* [5] developed Blnet, a hybrid architecture that combines ResNet, GoogleNet, and self-attention, which achieved an accuracy of 72.72% and improved sensitivity. However, the complexity of the model architecture and high computational demands may limit its application in clinical settings with limited resources. Compared to previous approaches, the method proposed in this study shows a balance between accuracy and practicality. By utilizing deep feature extraction from pre-trained CNNs with conventional machine learning classification algorithms, this approach achieves an accuracy of 73.58% with relatively low implementation complexity, providing greater potential for adaptation to real-world medical applications.

Table 7. Performance comparison of previous research

Ref	Method	Accuracy (%)
Demir *et al.* [4]	Transfer learning with a CNN model and a SoftMax classifier	65.50
		63.09
Yang *et al.* [5]	BInet with STFT+Wavelet transform	72.72
Proposed research	VGG-16+SVM linear	70.60
	ResNet-50+SVM linear	64.93
	DenseNet-121+SVM linear	73.58

In future research, feature selection techniques should be applied to reduce data dimensions and identify the most informative features, as the current study still relies on all features without evaluating their importance. Interpretability tools such as SHAP can also be considered to better understand the contribution of each feature. Additionally, performance improvements can be achieved by exploring alternative CNN architectures to generate more comprehensive feature representations and optimizing hyperparameters in classification models, such as SVM and random forest.

4. CONCLUSION

This research aims to determine the performance of different machine-learning algorithms for respiratory disease classification based on lung sound audio with deep features extracted from a CNN architecture specifically VGG-16, ResNet-50, and DenseNet-121. This research also applies 10-fold cross-validation to reduce the possibility of overfitting. The results show that the three CNN architectures perform best with the linear kernel SVM, VGG-16 achieved 70.63% accuracy and 70.51% F1-score, ResNet-50 achieved 64.93% accuracy and 64.73% F1-score, and DenseNet-121 achieved 73.58% accuracy and 73.37% F1-score.

Although the results obtained show promise, several limitations need to be considered. The overall accuracy has not reached the threshold of >80%, which is thought to be caused by the large number of features and the possibility of noise or irrelevant information from the deep feature extraction results. This condition suggests that further optimization strategies are necessary, although the CNN-based deep feature

extraction approach holds promise. Future research should explore other deep feature techniques, optimize hyperparameters, and apply feature selection and feature importance analysis to reduce data dimensions, eliminate irrelevant attributes, and identify the most informative features, thereby improving both efficiency and predictive performance. Additionally, ablation analysis was not included in this study, which is advised for future research, as it could assist in elucidating the impact of each component.

The primary contribution of this study is a systematic evaluation of deep feature extraction using a CNN architecture and machine learning algorithms for classifying respiratory diseases based on lung sound signal spectrograms. These findings can serve as a basis for developing medical audio signal-based diagnostic support systems. Furthermore, the suggested framework exhibits promise for incorporation into digital telemedicine systems, where respiratory sound analysis may facilitate remote diagnosis and ongoing patient monitoring via digital stethoscope recordings or smartphone-based apps. Access to respiratory healthcare services may be made easier by such integration, especially in places with limited resources or that are remote.

ACKNOWLEDGMENTS

This work was supported by the academic collaboration and institutional assistance of Lambung Mangkurat University and Kanazawa University. The author sincerely appreciates their continuous encouragement and support.

FUNDING INFORMATION

This research was financially supported by Lambung Mangkurat University Research Grant, year 2024, with grant number 1374. 68/UN8.2/PG/2024.

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
Elvina Nur Hana		✓	✓			✓		✓	✓					
Mohammad Reza Faisal	✓	✓							✓	✓		✓		✓
Dwi Kartini	✓			✓	✓				✓			✓		
Muhammad Itqan Mazdadi		✓				✓				✓	✓	✓		
Setyo Wahyu Saputro		✓	✓			✓				✓		✓		
Fatma Indriani	✓			✓	✓				✓					
Kenji Satou	✓					✓	✓			✓				

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 dataset used in this study is publicly available and can be accessed through the following link: https://www.kaggle.com/datasets/vbookshelf/respiratory-sound-database.

REFERENCES

[1] World Health Organization, "The top 10 causes of death." [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death. (Accessed: Jan. 20, 2025).

[2] B. M. Rocha, D. Pessoa, A. Marques, P. Carvalho, and R. P. Paiva, "Automatic Classification of Adventitious Respiratory Sounds: A (Un)Solved Problem?," *Sensors*, vol. 21, no. 1, pp. 1-19, Dec. 2020, doi: 10.3390/s21010057.
[3] B. A. Tessema, H. Nemomssa, and G. L. Simegn, "Acquisition and Classification of Lung Sounds for Improving the Efficacy of Auscultation Diagnosis of Pulmonary Diseases," *Medical Devices: Evidence and Research*, vol. 15, pp. 89–102, Apr. 2022, doi: 10.2147/MDER.S362407.
[4] F. Demir, A. Sengur, and V. Bajaj, "Convolutional neural networks based efficient approach for classification of lung diseases," *Health Information Science and Systems*, vol. 8, no. 1, pp. 1-8, Dec. 2020, doi: 10.1007/s13755-019-0091-3.
[5] R. Yang, K. Lv, Y. Huang, M. Sun, J. Li, and J. Yang, "Respiratory Sound Classification by Applying Deep Neural Network with a Blocking Variable," *Applied Sciences*, vol. 13, no. 12, 2023, doi: 10.3390/app13126956.
[6] S. W. Ali *et al.*, "Towards the Development of the Clinical Decision Support System for the Identification of Respiration Diseases via Lung Sound Classification Using 1D-CNN," *Sensors*, vol. 24, no. 21, pp. 1–16, 2024, doi: 10.3390/s24216887.
[7] H. Kumar and M. Aruldoss, "Gated Cross-Modal Fusion Mechanism for Audio-Video-based Emotion Recognition," *Engineering, Technology & Applied Science Research*, vol. 15, no. 2, pp. 20835–20841, Apr. 2025, doi: 10.48084/etasr.9430.
[8] J. P. Garcia-Mendez *et al.*, "Machine Learning for Automated Classification of Abnormal Lung Sounds Obtained from Public Databases: A Systematic Review," *Bioengineering*, vol. 10, no. 10, pp. 1-19, Oct. 2023, doi: 10.3390/bioengineering10101155.
[9] N. Gaffoor and S. Soomro, "Skin Disease Detection and Classification Using ResNet-50 and Support Vector Machine: An Effective Approach for Dermatological Diagnosis," in *2023 IEEE International Conference on Internet of Things and Intelligence Systems (IoTaIS)*, Bali, Indonesia, Nov. 2023, pp. 140–145, doi: 10.1109/IoTaIS60147.2023.10346059.
[10] K. V. Priya and J. D. Peter, "A federated approach for detecting the chest diseases using DenseNet for multi-label classification," *Complex and Intelligent Systems*, vol. 8, no. 4, pp. 3121–3129, 2022, doi: 10.1007/s40747-021-00474-y.
[11] A. Younis, L. Qiang, C. O. Nyatega, M. J. Adamu, and H. B. Kawuwa, "Brain Tumor Analysis Using Deep Learning and VGG-16 Ensembling Learning Approaches," *Applied Sciences*, vol. 12, no. 14, pp. 1-20, 2022, doi: 10.3390/app12147282.
[12] A. Rayan and A. S. Alaerjan, "An improved crow search optimization with Bi-LSTM model for identification and classification of COVID-19 infection from chest X-Ray images," *Alexandria Engineering Journal*, vol. 76, pp. 787–798, 2023, doi: 10.1016/j.aej.2023.06.052.
[13] S. K, D. Koppad, P. Kumar, N. A. Kantikar, and S. Ramesh, "Multi-Task Learning for Lung sound & Lung disease classification," *arXiv preprint*, Apr. 2024, doi: 10.48550/arXiv.2404.03908.
[14] A. Rahman, S. S. Rizvi, A. Khan, A. A. Abbasi, S. U. Khan, and T. S. Chung, "Parkinson's disease diagnosis in cepstral domain using MFCC and dimensionality reduction with SVM classifier," *Mobile Information Systems*, pp. 1-10, 2021, doi: 10.1155/2021/8822069.
[15] E. ŞAHiN, N. N. Arslan, and D. Özdemir, "Unlocking the black box: an in-depth review on interpretability, explainability, and reliability in deep learning," *Neural Computing and Applications*, vol. 37, no. 2, pp. 859–965, Jan. 2025, doi: 10.1007/s00521-024-10437-2.
[16] A. Wei, K. Yu, F. Dai, F. Gu, W. Zhang, and Y. Liu, "Application of Tree-Based Ensemble Models to Landslide Susceptibility Mapping: A Comparative Study," *Sustainability*, vol. 14, no. 10, p. 6330, May 2022, doi: 10.3390/su14106330.
[17] S. Naiem, A. E. Khedr, A. M. Idrees, and M. I. Marie, "Enhancing the Efficiency of Gaussian Naïve Bayes Machine Learning Classifier in the Detection of DDOS in Cloud Computing," *IEEE Access*, vol. 11, pp. 124597–124608, 2023, doi: 10.1109/ACCESS.2023.3328951.
[18] D. Hsu, V. Muthukumar, and J. Xu, "On the proliferation of support vectors in high dimensions," *Proceedings of Machine Learning Research*, vol. 130, pp. 91–99, 2021.
[19] R. K. Halder, M. N. Uddin, M. A. Uddin, S. Aryal, and A. Khraisat, "Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications," *Journal of Big Data*, vol. 11, no. 1, p. 113, Aug. 2024, doi: 10.1186/s40537-024-00973-y.
[20] A. Bricout *et al.*, "Bee Together: Joining Bee Audio Datasets for Hive Extrapolation in AI-Based Monitoring," *Sensors*, vol. 24, no. 18, p. 6067, Sep. 2024, doi: 10.3390/s24186067.
[21] P. G. Anjitha and E. D. Dileesh, "Epilepsy Detection using Spectogram data and Convolutional Neural Networks," in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Delhi, India, 2023, pp. 1–6, doi: 10.1109/ICCCNT56998.2023.10307755.
[22] A. O. Almagrabi, "A Deep CNN-LSTM-Based Feature Extraction for Cyber-Physical System Monitoring," *Computers, Materials and Continua*, vol. 76, no. 2, pp. 2079–2093, 2023, doi: 10.32604/cmc.2023.039683.
[23] Y. Liu, H. Pu, and D. W. Sun, "Efficient extraction of deep image features using convolutional neural network (CNN) for applications in detecting and analysing complex food matrices," *Trends in Food Science and Technology*, vol. 113, Oct. 2020, pp. 193–204, 2021, doi: 10.1016/j.tifs.2021.04.042.
[24] 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, 2023, doi: 10.1007/s11831-023-09899-9.
[25] B. Paneru, B. Paneru, and K. B. Shah, "Analysis of Convolutional Neural Network-based Image Classifications: A Multi-Featured Application for Rice Leaf Disease Prediction and Recommendations for Farmers," *Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 6, no. 3, Aug. 2024, doi: 10.35882/ijeeemi.v6i3.3.
[26] K. Palanisamy, D. Singhania, and A. Yao, "Rethinking CNN Models for Audio Classification," *arXiv preprint*, 2020, doi: 10.48550/arXiv.2007.11154.
[27] N. Zakaria, F. Mohamed, R. Abdelghani, and K. Sundaraj, "Three ResNet Deep Learning Architectures Applied in Pulmonary Pathologies Classification," in *2021 International Conference on Artificial Intelligence for Cyber Security Systems and Privacy (AI-CSP)*, El Oued, Algeria, Nov. 2021, pp. 1–8, doi: 10.1109/AI-CSP52968.2021.9671211.
[28] G. Sriram, T. R. G. Babu, R. Praveena, and J. V. Anand, "Classification of Leukemia and Leukemoid Using VGG-16 Convolutional Neural Network Architecture," *Molecular & Cellular Biomechanics*, vol. 19, no. 1, pp. 29–40, 2022, doi: 10.32604/mcb.2022.016966.
[29] S. A. Albelwi, "Deep Architecture based on DenseNet-121 Model for Weather Image Recognition," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 10, pp. 559–565, 2022, doi: 10.14569/IJACSA.2022.0131065.
[30] A. P. Ratnasari, "Performance of Random Oversampling, Random Undersampling, and SMOTE-NC Methods in Handling Imbalanced Class in Classification Models," *International Journal of Scientific Research and Management (IJSRM)*, vol. 12, no. 04, pp. 494–501, Apr. 2024, doi: 10.18535/ijsrm/v12i04.m03.
[31] T. Ait tchakoucht, B. Elkari, Y. Chaibi, and T. Kousksou, "Random forest with feature selection and K-fold cross validation for predicting the electrical and thermal efficiencies of air based photovoltaic-thermal systems," *Energy Reports*, vol. 12, pp. 988–

999, 2024, doi: 10.1016/j.egyr.2024.07.002.
[32] S. M. Malakouti, M. B. Menhaj, and A. A. Suratgar, "The usage of 10-fold cross-validation and grid search to enhance ML methods performance in solar farm power generation prediction," *Cleaner Engineering and Technology*, vol. 15, pp. 1-7, Jul. 2023, doi: 10.1016/j.clet.2023.100664.
[33] X. Tang, F. Li, Z. Cao, Q. Yu, and Y. Gong, "Optimising Random Forest Machine Learning Algorithms for User VR Experience Prediction Based on Iterative Local Search-Sparrow Search Algorithm," in *2024 6th International Conference on Communications, Information System and Computer Engineering (CISCE)*, Guangzhou, China, Jun. 2024, doi: 10.1109/CISCE62493.2024.10653373.
[34] M. A. Almaiah *et al.*, "Performance Investigation of Principal Component Analysis for Intrusion Detection System Using Different Support Vector Machine Kernels," *Electronics*, vol. 11, no. 21, pp. 1-16, Nov. 2022, doi: 10.3390/electronics11213571.
[35] O. P. Barus, Romindo, and Jefri Junifer Pangaribuan, "Classification of Hearing Loss Degrees with Naive Bayes Algorithm," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 4, pp. 751–757, Aug. 2023, doi: 10.29207/resti.v7i4.4683.
[36] V. B. Shtino and M. Muça, "Comparative Study of K-NN, Naive Bayes and SVM for Face Expression Classification Techniques," *Balkan Journal of Interdisciplinary Research*, vol. 9, no. 3, pp. 23–32, Dec. 2023, doi: 10.2478/bjir-2023-0015.
[37] F. J. M. Shamrat, S. Azam, A. Karim, K. Ahmed, F. M. Bui, and F. De Boer, "High-precision multiclass classification of lung disease through customized MobileNetV2 from chest X-ray images," *Computers in Biology and Medicine*, vol. 155, pp. 1-14, Mar. 2023, doi: 10.1016/j.compbiomed.2023.106646.
[38] M. S. H. Talukder *et al.*, "JutePestDetect: An intelligent approach for jute pest identification using fine-tuned transfer learning," *Smart Agricultural Technology*, vol. 5, pp. 1-13, Oct. 2023, doi: 10.1016/j.atech.2023.100279.
[39] J. S. Aguilar-Ruiz, "Beyond the ROC Curve: The IMCP Curve," *Analytics*, vol. 3, no. 2, pp. 221–224, May 2024, doi: 10.3390/analytics3020012.
[40] M. C. Hinojosa Lee, J. Braet, and J. Springael, "Performance Metrics for Multilabel Emotion Classification: Comparing Micro, Macro, and Weighted F1-Scores," *Applied Sciences*, vol. 14, no. 21, pp. 1-21, Oct. 2024, doi: 10.3390/app14219863.
[41] A. Ibias, K. Capala, V. R. Varma, A. Drozdz, and J. Sousa, "Improving Noise Robustness through Abstractions and its Impact on Machine Learning," *arXiv preprint*, Jun. 2024, doi: 10.48550/arXiv.2406.08428.

BIOGRAPHIES OF AUTHORS

Elvina Nur Hana was born in South Kalimantan's Banjarbaru City. She has been enrolled in Lambung Mangkurat University's Computer Science program since 2021. She is now working on her undergraduate degree and is interested in data science research. She has finished several data processing-related projects. She is dedicated to advancing research and expanding her knowledge and expertise in information technology. She can be contacted at email: 2111016220002@mhs.ulm.ac.id.

Mohammad Reza Faisal after he completing high school, he attended Pasundan University in 1995 to complete his undergraduate studies in the informatics department. In 1997, he majored in physics at Bandung Institute of Technology. After finishing his bachelor's degree, he obtained expertise as a software developer and information technology trainer. He has been teaching computer science at Universitas Lambung Mangkurat since 2008. In 2010, he completed his master's degree in informatics at Bandung Institute of Technology. He continued his studies in 2015 by attending Kanazawa University in Japan to pursue a doctorate in bioinformatics. He is still employed at Universitas Lambung Mangakurat as a computer science lecturer. He is interested in bioinformatics, software engineering, and data science. He can be contacted at email: reza.faisal@ulm.ac.id.

Dwi Kartini obtained her bachelor's and master's degrees in computer science from Putra Indonesia University "YPTK" in Padang, Indonesia. She is interested in data mining and artificial intelligence applications. She is an assistant professor at Lambung Mangkurat University in Banjarbaru. Indonesia. in the Department of Computer Science. Faculty of Mathematics and Natural Sciences. She can be contacted at email: dwikartini@ulm.ac.id.

Muhammad Itqan Mazdadi is a Lambung Mangkurat University's Department of Computer Science lecturer. Computer networking and data science are his main areas of interest. He earned his bachelor's degree in computer science from Lambung Mangkurat University in 2013 before starting as a lecturer. After that. he earned his master's degree from Islamic Indonesia University in Yogyakarta's Department of Informatics. He is currently Lambung Mangkurat University's Department of Computer Science Secretary. He is committed to creating a creative and cooperative learning atmosphere that inspires students to pursue and succeed in computer science. He can be contacted at email: mazdadi@ulm.ac.id.

Setyo Wahyu Saputro lectures at Lambung Mangkurat University in Banjarbaru in the Department of Computer Science, Faculty of Mathematics and Natural Science, Lambung Mangkurat University awarded him a bachelor's degree in computer science in 2011. Moreover, STMIK Amikom University awarded him a master's degree in informatics in 2016. Since 2017, he has worked as a consultant and information technology practitioner, managing and analyzing systems for several government and private projects in South Kalimantan. His areas of interest include artificial intelligence applications, software engineering, and human-computer interaction. He can be contacted at email: setyo.saputro@ulm.ac.id.

Fatma Indriani is lectures at Lambung Mangkurat University in Indonesia, where she teaches computer science. She graduated from Monash University in Australia with a master's degree in computer science in 2012, and Kanazawa University in Japan awarded her a doctorate in bioinformatics in 2022. Data preparation, natural language processing, bioinformatics, image processing, and machine learning applications in the biomedical sciences and healthcare are the main areas of her research. She has worked on text mining for social media and educational data analysis, deep learning for picture classification, and feature engineering approaches for enhancing predictive models. She can be contacted at email: f.indriani@ulm.ac.id.

Kenji Satou received his bachelor's and master's degrees in Computer Science and Communication Engineering from Kyushu University in 1987 and 1989, respectively. He earned his Doctorate in Engineering and started his academic career as a Research Associate at Kyushu University and then at the Human Genome Center at the University of Tokyo. From 1998 to 2007, he served as an Associate Professor at the Japan Advanced Institute of Science and Technology before joining Kanazawa University in 2007. He is a professor at the Institute of Philosophy in Interdisciplinary Sciences, Kanazawa University. His research interests include intelligent informatics and bioinformatics, particularly in developing deductive database systems for protein structure analysis. He is active in various scientific organizations. including the Information Processing Society of Japan and the Japanese Society for Bioinformatics. He can be contacted at email: ken@t.kanazawa-u.ac.jp.