Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 8, No. 2, April 2026, pp: 504-516 e-ISSN: 2656-8632

RESEARCH ARTICLE OPEN ACCESS

MK–TripNet: A Deep Learning Framework for Real-Time Multi-Class Lung Sound Classification

Widya Surya Erini[1], Gracia Putri Thomas[1], Giulia Salzano Badia[1], Arief Rahadian[2], Sofyan Budi Raharjo[3], Sari Ayu Wulandari[1]

[1] Department of Biomedical Engineering, Universitas Dian Nuswantoro, Semarang, Central Java, Indonesia
[2] Department of Medicine, Universitas Dian Nuswantoro, Semarang, Central Java, Indonesia
[3] Department of Internal Medicine, Dr. Kariadi General Hospital, Semarang, Central Java, Indonesia

Corresponding author: Sari Ayu Wulandari (e-mail: sari.wulandari@dsn.dinus.ac.id), **Author(s) Email**: Widya Surya Erini (e-mail: widyasuryaelini@gmail.com), Gracia Putri Thomas (e-mail: graciaputrithomas@gmail.com), Giulia Salzano Badia (e-mail: giuliabadia29@gmail.com), Arief Rahadian (e-mail: dr.ariefrahadian@gmail.com), Sofyan Budi Raharjo (e-mail: so_arjopulmo@yahoo.com)

Abstract **Respiratory diseases such as asthma, pneumonia, and Chronic Obstructive Pulmonary Disease (COPD) remain major global health challenges, particularly in resource-limited settings where access to pulmonary specialists and early diagnostic tools is limited. Automatic lung sound classifications have emerged as a promising non-invasive screening approach; however, existing methods often rely on single-scale feature extraction, conventional loss functions, and offline analysis, which limit their discriminative capability and real-time applicability. The aim of this study is to develop and evaluate a deep learning framework for real-time multi-class lung sound classifications that improves discriminative representation and temporal sensitivity. To address limitations, this study proposes MK-TripNet, a novel deep learning architecture designed to integrate multi-scale feature extraction, discriminative embedding learning, and real-time inference within a unified framework. The main contribution of this work is the unified integration of a Multi-Kernel convolutional architecture, Triplet Loss-based embedding learning, and Sliding Window segmentation within a single end-to-end framework, enabling accurate segment-level lung sound classifications in real-time scenarios. Unlike prior approaches, the proposed method simultaneously captures fine-grained temporal patterns and broader spectral characteristics while explicitly maximizing inter-class separability in the embedding space. The proposed model was evaluated using a newly constructed dataset comprising 1,409 lung sound segments obtained from primary digital stethoscope recordings and publicly available respiratory sound databases. Experimental results demonstrate that MK-TripNet consistently outperforms several strong baseline models, including CNN-BiGRU, CNN-BiGRU-UMAP, and VGGish-Triplet, achieving an accuracy of 89.1%, an F1-score of 0.89, and a recall of 0.88. Ablation studies further confirm that the combined use of Multi-Kernel convolution, Triplet Loss, and Sliding Window segmentation yields the most robust and generalizable performances. These findings highlight the clinical potential of MK-TripNet for real-time digital auscultation and point-of-care respiratory screening, particularly in resource-limited and telemedicine settings.**

Keywords **Multi Kernel; Triplet Loss; Sliding Window; CNN; MFCC**

I. Introduction

Respiratory diseases such as COPD, asthma, chronic bronchitis, lung cancer, pneumonia, tuberculosis, and COVID-19 remain the leading causes of death and disability in the world [1], [2], with COPD representing a significant global health challenge driven by smoking as well as environmental, occupational and genetic factors [3], as reported in recent global burden trends of COPD [4]. Although there has been extensive research on the classification of breathing sounds using machine learning, traditional approaches remain dominant, relying solely on features such as MFCC or spectrograms and general architectures (CNN, CNN-LSTM, CNN-BiGRU) without advanced representation techniques [5]. Several studies have been conducted with various approaches. For example, CNN-BiGRU was applied on synthetic data without a specific loss function [6]. Another work focused mainly on dataset creation without integrated classification [7]. UMAP was used only for visualization [8]. DLCNN achieved high accuracy (98.85%) but was limited by the representativeness of the data [9]. A deep CNN was applied to dependent noise synthesis [10]. CNN-LSTM was employed, but with a limited dataset [11]. Lung and

Manuscript received 13 December 2025; Revised 20 February 2026; Accepted 15 March 2026; Available online 30 March 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1403

tracheal sounds were analyzed using MFCC, but constrained by acoustic differences [12]. Recent work employed a CNN leveraging mel spectrograms to identify adventitious lung sounds, while Bardou et al [13] combined MFCC and traditional ML features with CNN-based classifiers. Multi-Kernel CNN and TMK-CNN were also developed, but with an unbalanced dataset and no internal validation. In summary, many studies report high accuracy with deep CNN or hybrid models, but often rely on unbalanced datasets, lack external validation, or depend solely on manual augmentation [11].TMK-CNN with Multi-Kernel and Triplet Loss remains limited to a single dataset without real-time segmentation. Unsupervised methods such as UMAP also lack a discriminative training objective like Triplet Loss. While previous studies have explored Multi-Kernel convolutional, Triplet Loss, or temporal segmentation independently or in partial combinations, a unified framework that jointly integrates these components for real-time multi-class lung sound classification remains insufficiently investigated. Specifically, MK-TripNet addresses data imbalance by constructing a composite dataset, improves class separability via metric learning, and enables real-time inference through Sliding Window segmentation. Most CNN studies still rely on Single Kernel, cross-entropy, or focal loss, without real-time inference [14]. Some recent works have attempted related approaches. For example, feature fusion from several pre-trained CNNs enhances discriminative representations from lung sound spectrograms [15]. A lightweight CNN with multi-feature integration achieves strong performance in lung sound classification [16]. Transfer learning with multi-input CNN models is used for crackle detection across differing recording setups [17]. The use of multi-channel recordings combined with CNN-LSTM improves sensitivity and specificity compared with fewer channels [18], and models that use multi-channel spectrograms with attention and augmentation address data imbalance [19]. Despite these advances, important gaps remain. Therefore, this paper proposes MK-TripNet, a Multi-Kernel CNN with Triplet Loss that combines multi-scale temporal representation, discriminative embedding learning, and real-time inference on a newly balanced dataset. Some studies employed multi time scale or hybrid features but without Triplet Loss [20], while Multi-Kernel convolution has shown improved temporal and spectral pattern extraction in lung sounds [21].

MK-TripNet (Multi-Kernel CNN with Triplet Loss) is a neural network-based model for detecting pulmonary diseases (PDs) through real-time analysis of lung sound (LS) signals. Real-time inference refers to the sequential processing of overlapping audio segments with minimal delay, enabling continuous analysis during auscultation rather than post-hoc classification. The model is specifically designed to capture a broader range of acoustic features by utilizing Multi-Kernel convolution blocks at various scales, thereby improving classification accuracy [14], [22]. In addition, Triplet Loss, which has been shown to be effective in producing more distinct feature representations across classes [23], [24], is employed to generate more discriminative embeddings. By simultaneously processing anchor, positive, and negative inputs, MK-TripNet can distinguish between similar sounds such as wheezing and crepitus, and subsequently classify them into four categories: bronchus, crepitus, wheezing, or normal through a softmax layer [25], [26], [27], [28], [29].

The primary objective of this study is to develop and validate MK-TripNet, a deep learning model designed for Multi-Class lung sound classification in real-time settings. The contributions of this work are 1) we introduce a new hybrid dataset that combines both primary recordings and secondary public sources, preprocessed and segmented using overlapping windows to enable real-time classification consistent with clinical auscultation practices [30]. 2) We demonstrate that MK-TripNet achieves superior performance compared to baseline models such as CNN-BiGRU and CNN-BiGRU+UMAP in segment-based multi-class classification, confirming its effectiveness for distinguishing between normal, wheeze, crackle, and bronchial sounds [31]. 3) We perform a comprehensive ablation study to examine the influence of Kernel size, embedding dimension, and number of filters, thereby proving the robustness and generalization capability of MK-TripNet across different configurations and datasets [32]. 4) By incorporating Triplet Loss embedding, our model enhances discriminative feature learning for respiratory audio, supporting more reliable differentiation of acoustically similar lung sound patterns. Prior works have demonstrated the effectiveness of Triplet Loss and Siamese-based frameworks in improving inter-class separability and representation learning. Specifically, a triple-network architecture outperformed the ICBHI baseline, improving the overall score by ~7.0% [23]. A Multi Triplet Loss improved the inter-class separability of data representations in the embedding space [33]. A Siamese network with Triple Loss was applied to evaluate synthetic respiratory sounds for privacy-preserving dataset generation [34]. Triple Loss was shown to narrow intra-class distances and widen inter-class distances, enhancing cross-domain generalization [35]. Furthermore, a triplet Siamese framework was successfully applied to medical image classification, demonstrating its effectiveness in discriminating between clinically similar categories [36]. This study is structured as follows: Section II related work on lung sound classification, embedding

Manuscript received 13 December 2025; Revised 20 February 2026; Accepted 15 March 2026; Available online 30 March 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1403

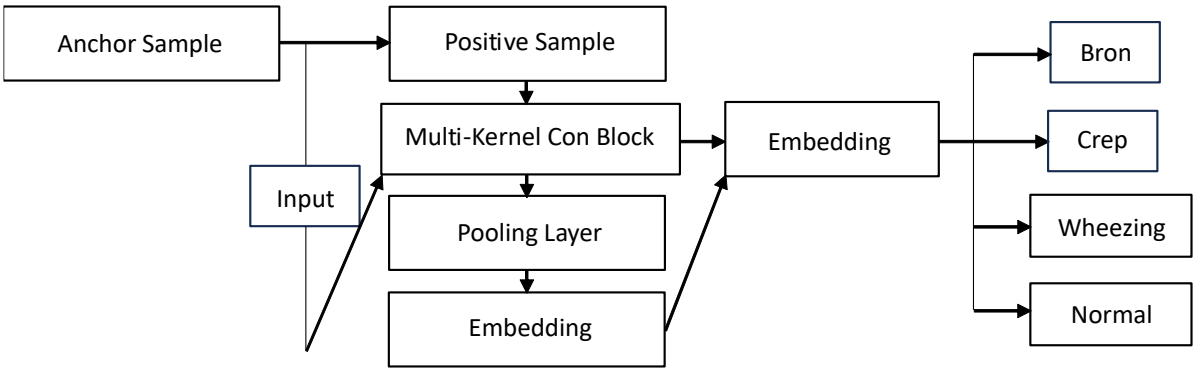

Fig. 1. Architecture of Multi-Kernel CNN Using Triplet Loss Function for Classification Tasks

learning techniques, Multi-Kernel architecture, and real-time audio inference. Section III presents the detailed architecture of MK-TripNet, including its Multi-Kernel convolution blocks, Triplet Loss embedding head, and Sliding Window inference mechanism. Section IV describes the dataset used, preprocessing procedures, training protocol, and evaluation metrics. Section V reports the experimental results, ablation studies, cross-dataset validations, and comparisons with baseline methods. Finally, section VI concludes the work and discusses future directions toward deployment in clinical environments.

II. Method

This study aims to examine whether there is a significant difference in accuracy between the Multi-Kernel CNN Architecture with Triplet Loss and other methods. In this study, we propose MK-TripNet, a deep learning architecture for multi-class classification of lung audio signals. The model utilizes a Multi-Kernel Convolutional Block that processes inputs (anchor, positive, negative) in parallel with kernels of size 1x1, 3x3, and 5x5 to capture temporal and spectral patterns of various scales. Each kernel branch is followed by normalization and ReLU activation function, with the number of output channels set to 64, 128, and 256, respectively. The resulting feature maps from all branches are concatenated along the channel dimension and subsequently passed to a pooling layer for dimensionality reduction. The merged features are then processed through a fully connected layer to obtain a compact embedding representation. This structure uses a triplet sampling approach during training, which involves three types of audio input: anchor (x^a), positive (x^p), and negative (x^n), where anchor and positive are from the same class, while negative is from a different class, as illustrated in Fig. 1. The goal is for the model to be able to distinguish similar and dissimilar sound features in the latent representation space. Each input is processed through parallel convolution results are then merged, normalized using batch normalization, and activated with the ReLU function Eq. (1) [23].

$$Conv_1(x) = W_1 \times x + b_1$$
$$Conv_3(x) = W_3 \times x + b_3$$
$$Conv_5(x) = W_5 \times x + b_5 \quad (1)$$

Where x is the audio signal input feature map, W_k and b_k are the weight matrix bias associated with the convolution kernel of size $k \in \{1,3,5\}$, and x is the convolution operation. Then, the three outputs are combined in Eq. (2) [14]:

$$MKConv(x) = ReLU(BN(|Conv_1(x), Conv_3(x), Conv_5(x)|)) \quad (2)$$

Where $BN(\cdot)$ is batch normalization to stabilize the training process, $ReLU(\cdot)$ is the rectified linear unit activation function, and $(\cdot)$ denotes the concatenation of feature maps along the channel dimension. After convolution, max pooling is used to reduce the spatial size while retaining key features, Eq. (3) [14]:

$$Pooled(x) = MaxKernel\ (MKConv(x)) \quad (3)$$

Where $MaxKernel(\cdot)$ denotes the max pooling operation with a predefined kernel size and stride. This stage is important for selecting significant features and simplifying calculations. The pooling results are condensed into a one-dimensional vector and passed to the dense layer to form a fixed- dimensional embedding $\mathbb{R}^d$ Eq. (4) [14]:

$$f(x) = FC\ \left(Flatten(x)\right) \epsilon \mathbb{R}^d \quad (4)$$

The trained d dimensional embedding is represented by $f(x) \in \mathbb{R}^d$, where $FC(\cdot)$ denotes a fully connected layer, and $Flatten(\cdot)$ converts the pooled feature map into a one-dimensional vector. This vector represents the hidden features of the audio signal for discriminative learning. Triplet Loss is applied with anchor, positive, and negative data from the training data, which serves to minimize the distance between similar embeddings and widen the distance between different classes, Eq. (5) [23]:

$$\mathcal{L}_{triplet} = \max\left(0, ||f(x^a) - f(x^p)||^2 - ||f(x^a) - f(x^n)||^2 + a\right) \quad (5)$$

Manuscript received 13 December 2025; Revised 20 February 2026; Accepted 15 March 2026; Available online 30 March 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1403

Where a is the margin parameter, which is empirically selected as 0.2 to balance intra-class density with inter-class separation, and $x^a, x^p,$ and x^n represent the anchor, positive, and negative samples, respectively. To balance intra-class density and inter-class separation, the margin parameter a is empirically set to 0.2 based on preliminary experiments. In this case, the embedding representation of the input is denoted by $F(x)$. Four categories, bronchial, crepitus, wheezing, and normal, are predicted by SoftMax classification using only the anchor branch during inference. Additionally, the model uses Sliding Window segmentation with a window length of 1 second and a step of 0.5 seconds to enable real-time detection of abnormal breathing patterns and segment-level categorization, Eq. (6) [23]:

$$\hat{y} = softmax(Wf(xa) + b) \quad (6)$$

This layer generates probabilities for each category: bronchial, crepitus, wheezing, and normal. The MK-TripNet method is summarized in Algorithm 1.

Algorithm 1. Multi-Kernel CNN with Triplet Loss (MK-TripNet) for Lung Sound Classification

Input: anchor (x_a), positive (x_p), and negative (x_n)
Output: $\hat{y}$

1. Take 1 anchor sample (e.g. "bron" sound)
2. Take 1 positive sample (another sound than "bron")
3. Take 1 negative sample (sound from another class, e.g "normal")
4. Process all three → extract MFCC →CNN →Vector embedding
5. Calculate Triplet Loss
6. Backpropagation → model parameter optimization
7. Repeat for several epochs

A. Data Collection

Fig. 2 Modified acoustic stethoscope setup for data collection. A sensitive condenser microphone setup, as demonstrated in AR-based auscultation systems, digital stethoscope recordings on anterior and posterior chest walls, and standardized manikin recordings, is integrated into the stethoscope diaphragm to capture breath sounds in standard auscultation positions, which are digitally recorded for manual labelling and machine learning analysis [37], [38].

Fig. 2. Overview of Materials and Equipment Used for the Primary Dataset Collection

B. Dataset

Lung sound is a simple and non-invasive method that provides essential information for identifying respiratory disorders [39]. Respiratory sounds, produced during coughing and breathing, can be easily recorded with minimal invasiveness and analyzed using machine learning algorithms to identify patterns associated with specific diseases [40]. The dataset consists of two parts: main data from direct observations and supporting data from a public repository containing high-quality lung sound recordings from patients and healthy individuals with seven different conditions (asthma, heart failure, pneumonia, bronchitis, pleural effusion, pulmonary fibrosis, COPD), as well as normal sounds. The secondary dataset was obtained from a publicly available lung sound respiratory Kaggle, namely A Dataset of Lung Sounds, which provides labeled lung sound recordings collected from multiple clinical resources. A total of 336 audio files (.Wav) are labeled according to lung sound type (wheezing, crackle, bronchial, normal) and recorded in different thoracic regions with varying repetitions and frequency filters in Table 1. For real-time analysis, the dataset is modified using Sliding Window segmentation so that each file contains only a lung sound example, facilitating model training and evaluation. A total of 1.409 lung sound samples were obtained to train and test the classification model. The mode was then tested in real time using a Sliding Window for segment classification

Table 1. New Comprehensive Distribution of the Annotated Clinical Lung Sound Dataset

No	Type of Sound	Number of Secondary Dataset	Number of Units Dataset	Number of Primary Dataset
1	Normal	35	337	144
2	Crepitus	33	195	83
3	Bronchial	3	12	5
4	Wheeze	32	443	190
Total				1409

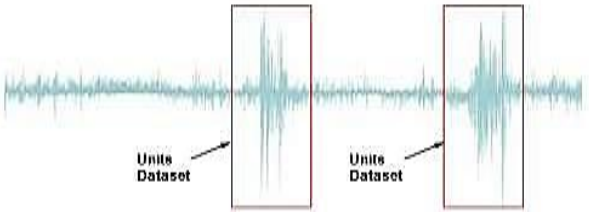

Fig. 3. Processing Units of the Dataset Derived from the Secondary Lung Sound Dataset

Manuscript received 13 December 2025; Revised 20 February 2026; Accepted 15 March 2026; Available online 30 March 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1403

and local sound pattern detection, in accordance with actual auscultation.

C. Experiment

This research involved four main experimental stages to thoroughly evaluate the performance of the MK-TripNet architecture using 1,409 segmented lung sound samples Fig. 3. First, multi-class classification was performed using several model variants: Baseline CNN with regular convolution, MK-CNN that adds a Multi-Kernel convolution block without Triplet Loss, then MK-CNN+Triplet with Triplet Loss applied, and finally MK- CNN+Triplet+Sliding Window, that forms the full MK-TripNet architecture. This evaluation aims to assess each component's contribution to the model's performance.

The second stage is an ablation study to evaluate the influence of various architectural parameters such as embedding dimension, number of convolutional channels, merging method, and dropout rate. For example, increasing the embedding dimension from 64 to 128, as well as the convolutional channels from 32-64-128 to 64-128-256, shows significant performance improvement, confirming MK-TripNet's flexibility and robustness to architectural configuration changes.

The third stage involved comparing MK-TripNet with other methods, such as MFCC- CNN, CNN-BiGRU, CNN-BiGRU-UMAP, CNN with Triplet Loss, and VGGish with Triplet Loss, which showed that the combination of Multi-Kernel convolution and Triplet Loss resulted in a more accurate latent representation. Finally, MK-TripNet was tested in a real-time scenario using a Sliding Window approach, where audio data was processed on a spectrogram, demonstrating the model's ability to identify crepitus sounds and normal sounds directly based on their frequency patterns.

D. Performance Testing

To evaluate the performance of the proposed multi-class classification model, confusion matrix analysis is used with four main metrics: accuracy, precision, recall, and F1-score. These metrics provide an overall picture of the classification quality, especially in the context of class imbalance. The confusion matrix helps identify the number of True Positives (TP), False Positives (FP), False Negatives (FN), and True Negatives (TN), which respectively reflect the model's success or error in classifying samples. The evaluation is based on the standard formulas of these four metrics. Accuracy is a measure of the number of correct predictions compared to the total sample, Eq. (7) [32]:

$$Accuracy = \frac{\Sigma_{i=1}^{C} TP_i}{Total\ Sample} \quad (7)$$

Where C is the total class. The accuracy for each category is described as shown in Eq. (8) [32]:

$$Precision_i = \frac{TP_1}{TP_i + FP_i} \quad (8)$$

The recall ability (or sensitivity level) for each group is defined by Eq. (9) [32]:

$$Recall_i = \frac{TP_i}{TP_i + FN_i} \quad (9)$$

The F1 value for each category is the harmonic mean between precision and recall, and is calculated using Eq. (10) [32]:

$$F1_i = \frac{2.precision_i.recall_i}{precision_i + recall_i} \quad (10)$$

The macro-average metric calculates the average score for each class without considering class size, whereas the weighted average accounts for the number of samples in each class. These two metrics are used to evaluate the modal's ability to fairly distinguish between the four classes, and are essential in comparing performance with recent approaches in similar studies.

III. Result

A. Ablation Study

The ablation analysis results in Table 2 show the contribution of each element, Multi-Kernel Block, Triplet Loss, and Sliding Window, to classification performance. Compared with Baseline-CNN without these three features, performance improved significantly when these elements were applied. Table 2 shows the contribution of Multi-Kernel Block, Triplet Loss, and Sliding Window to MK-TripNet. The Baseline-CNN without these three components achieved an accuracy of 78.2%, an F1-score of 0.74, and a recall of 0.71. Adding the Multi-

Table 2. Comprehensive Performance Analysis of the MK-TripNet Ablation Study Results

Model Variant	Multi-Kernel Block	Triplet Loss	Sliding Window	Accuracy (%)	F1-Score	Recall
Baseline CNN	-	-	-	78.2	074	0.71
MK-CNN (No Triplet)	✓	-	-	82.5	0.79	0.77
MK-CNN + Triplet Loss	✓	✓	-	86.4	0.84	0.83
MK-CNN + Triplet +Sliding Window	✓	✓	✓	87.1	0.86	0.85
MK-TripNet (Full)	✓	✓	✓	88.3	0.88	0.87

Manuscript received 13 December 2025; Revised 20 February 2026; Accepted 15 March 2026; Available online 30 March 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1403

Table 3. Comprehensive Performance Analysis of the Selected Hyperparameters within MK-TripNet Model

Variant	Embedding Dim	Conv Channels	Pooling	Dropout	Accuracy	F1-Score	Recall
MK-TripNet (Emb=64)	64	32-64-128	Max	-	85.4	0.82	0.81
MK-TripNet (Emb=128)	128	32-64-128	Max	-	88.3	0.88	0.87
MK-TripNet (Emb=256)	256	32-64-128	Max	-	88	0.87	0.86
MK-TripNet (Conv=32-64-128)	128	32-64-128	Max	-	88.3	0.88	0.87
MK-TripNet (Conv=64-128-256)	**128**	**64-128-256**	**Max**	**-**	**89.1**	**0.89**	**0.88**
MK-TripNet (Pool=Avg)	128	32-64-128	Avg	-	87.2	0.86	0.85
MK-TripNet (Pool=Max)	128	32-64-128	Max	-	88.3	0.88	0.87
MK-TripNet+Dropout (0.3)	128	32-64-128	Max	0.3	88.4	0.88	0.87
MK-TripNet+Dropout (0.5)	128	32-64-128	Max	0.5	87.9	0.87	0.86

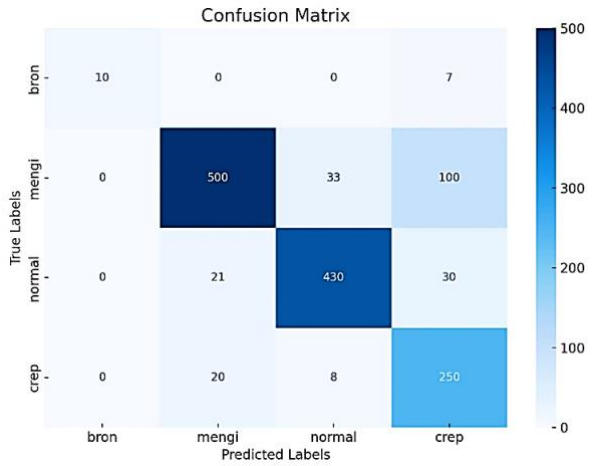

Fig. 4. Detailed Confusion Matrix of MK-TripNet Demonstrating Overall Best Performance Results

Kernel Block increased accuracy to 82.5%, F1-score to 0.79, and recall to 0.77, demonstrating the effectiveness of multi-scale convolution. Adding Triplet Loss improves the results to an accuracy of 86.4%, an F1-score of 0.84, and a recall of 0.83, proving better separation between classes. Sliding Window increases accuracy to 87.1% with an F1-score of 0.86 and a recall of 0.85, reinforcing temporal understanding. The complete MK- TripNet model achieves an accuracy of 88.3% an F1-score of 0.88, and a recall of 0.87, demonstrating the combined benefits of the three components. Hyperparameter tuning was also performed to further optimize performance.

B. Hyperparameter Study

A summary of MK-TripNet hyperparameter turning results is shown in Table 3. In this study, several key parameters were modified in a planned manner, such as embedding size, convolution channel depth, pooling type, and dropout rate, to measure their impact on model performance. An embedding size of 128 proved to be optimal (accuracy 88.3%, F1- score 0.88), while further increases did not provide significant gains. A convolution depth of deeper structures in capturing complex acoustic patterns. Max pooling performed slightly better than average pooling, and a moderate dropout rate (0.3) yielded the best generalization, whereas higher levels degraded performance. The optimal configuration consists of 128 embeddings, 64-

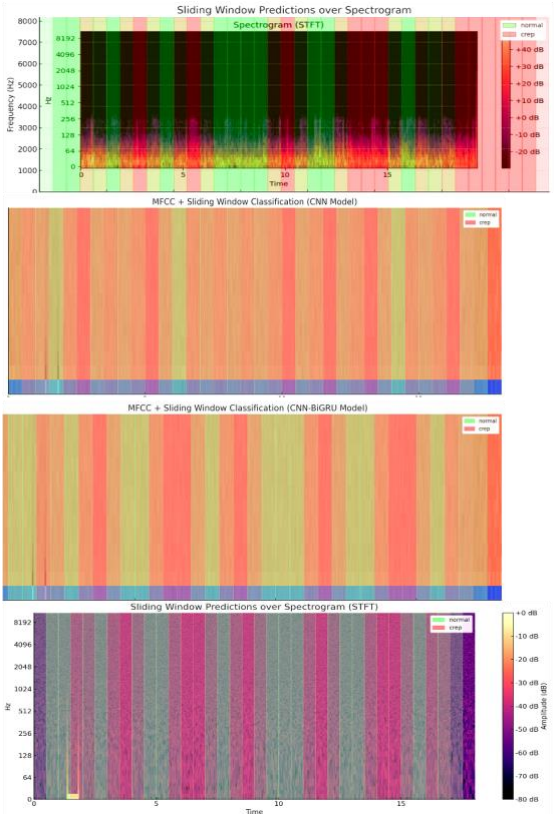

Fig. 5. Confusion Matrix Obtained from the Overall Best Performance of MK-TripNet

Manuscript received 13 December 2025; Revised 20 February 2026; Accepted 15 March 2026; Available online 30 March 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1403

128-256 convolutions, max pooling, and 0.3 dropout, which achieves the best balance between accuracy and generalization, as shown in the confusion matrix in Fig. 4.

This matrix illustrates the prediction accuracy of four lung sound categories: bronchus, wheeze, normal, and crepitus. The model successfully classified 10 out of 17 bronchial samples; the misclassification is only a shift towards crepitus, which clinically has a similar acoustic transient. For wheezing, 500 samples were correct, but 33 were incorrect as normal and 100 as crepitus, indicating confusion on subtle variations. The normal category had 430 correct predictions, with some incorrect as wheezing (21) and crepitus (30). Crepitus was recorded as 250 correct, but 20 incorrect as wheezing and 8 as normal, reflecting the similarity of temporal or spectral patterns.

C. Performance Comparison with Alternative Learning Strategies

Table 4 compares MK-TripNet with several baseline models based on different feature extraction and embedding, while Table 5 summarizes the quantitative comparison of accuracy, precision, recall, and F1-score between MK-TripNet and existing baseline models. MK-TripNet excelled across all metrics, achieving 89.1% accuracy, 84.5% precision, 0.88 recall, and 0.89 F1-score, proving the effectiveness of Multi-Kernel and Triplet Loss. The MFCC-CNN-BiGRU model achieved 83.5% accuracy and F1 of 0.82, but was limited by cross-entropy loss. CNN-BiGRU-UMAP was slightly better (85.1% accuracy and F1-scrore of 0.83) due to dimensionality reduction, though at the cost of increased complexity. In the CNN-BiGRU-UMAP baseline, UMAP was applied as a dimensionality reduction step prior to classification rather than solely for visualization, which differs from its use in related visualization-focused studies. CNN-Triplet Loss recorded 84.3% and F1 0.83, showing triplets are useful but not optimal without multi-scale processing. VGGish-Triplet Loss was lowest (82.7% and F1 0.79), indicating general learning transfer is less suitable for lung sounds without special adjustments.

D. Sliding Window Analysis

To illustrate the temporal dynamics captured by MK-TripNet, the audio sample “BP36 pneumonia, crep, P R M, 36, F”, from a 36-year-old female patient with pneumonia and crepitus in the right center posterior area of the chest, was used. Crepitation, a fine crackling sound that often occurs during inspiration, is common in pulmonary conditions such as pneumonia of fibrosis. Sliding Window analysis shows crepitant segments alternating with normal breath sounds, emphasizing the importance of temporal segmentation to detect subtle pathologic transitions in a single respiratory cycle. This example highlights the need for time-based local analysis so as not to overlook significant intra-sample variations, as visualized on the spectrogram in Fig. 5. This sample combines crepitus and normal breath sounds, challenging detailed acoustic segmentation. Manually labeled STT spectrograms (green: normal, red: crepitus) show the classification results of each model. CNN-BiGRU+MFCC in the middle panel predicts alternately but is unstable and overpredicts. In contrast, MK-

Table 4. Comprehensive Comparative Analysis of MK-TripNet Architecture with Various Alternative Learning

Model	Feature Extraction	Embedding Type	Convolution Type	Loss Function	Sliding Window	Real-Time Capable
MFCC+CNN	MFCC	-	Single Kernel Conv	Cross Entropy	-	-
CNN+BiGRU	MFCC	Bi-GRU (Recurrent)	Single Kernel Conv	Cross Entropy	-	-
CNN+Bi-GRU+UMAP	MFCC+Dimensionality Reduction (UMAP)	Bi-GRU+UMAP	Single Kernel Conv	Cross Entropy	-	-
MK-TripNet {Proposed Method)	MFCC	Triplet Loss Embedding	Multi Kernel Conv (1x1, 3x3, 5x5)	Triplet Loss+Cross Entropy	✓	✓

Table 5. Quantitative Comparison of the MK-TripNet Architecture with Alternative Learning Approaches

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
MK-TripNet {Proposed}	89.1	84.5	088	0.89
MFCC-CNN-BiGRU	83.5	81.2	0.82	0.82
CNN-BiGRU-UMAP	85.1	83.7	0.83	0.83
CNN-Triplet Loss	84.3	82.9	0.84	0.83
VGGish-Triplet Loss	82.7	80.1	0.79	0.79

Manuscript received 13 December 2025; Revised 20 February 2026; Accepted 15 March 2026; Available online 30 March 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1403

TripNet in the bottom panel provides smooth and consistent labeling, accurately detecting overlapping sound transitions.

The Sliding Window results in Fig. 5 **demonstrate** high prediction stability, which is further validated by the classification data where the model accurately predicted 500 wheezing samples and 430 normal samples. Despite the rapid transition within the respiratory cycle, MK-TripNet minimizes detection errors in challenging classes such as Bronchial, achieving an accuracy of 10 out of 17 samples, a performance significantly more stable than the baseline models. These advantages come from the Multi-Kernel convolution that captures patterns across multiple scales, the Triplet Loss that forms a discriminative embedding, as well as Sliding Windows on the feature map that preserve temporal and spatial context. while MFCC-based or recursive models without metric learning are prone to errors in complex acoustic areas. These findings confirm the importance of multi-scale integration, discriminative embedding, and effective temporal segmentation in MK-TripNet for real-time lung sound classification.

E. Discussion

A. Ablation Study

The aims of this study are to determine whether the application of various architectural elements (Multi-Kernel Block, Triplet Loss, and Sliding Window segmentation) separately and in combination produces significant variations in lung sound classification performance. by adding each component gradually to observe its impact on classification performance, ablation analysis compares the proposed MK-TripNet model with a simple CNN model. As each component is added to the base model, the experimental findings listed in Table 2 show a consistent improvement in classification accuracy. The addition of Multi-Kernel Block increases accuracy by 4.3% (82.5% vs. 78.2%) compared to the baseline CNN. When Triplet Loss is added, there is an additional accuracy improvement of 3.9% (86.4% vs. 82.5%). A smaller but stable improvement of 0.7% (87.1% vs. 86.4%) is achieved using Sliding Window segmentation. With an overall accuracy of 88.3%, the complete MK-TripNet model outperformed baseline CNN, Multi-Kernel only, and Multi-Kernel+Triplet by 10.1%, 5.8%, and 1.2%, respectively.

The observed performance improvement demonstrates how the non-stationary nature of lung sound signals and the MK-TripNet architecture design are mutually compatible. Short and long duration auditory patterns can be modeled simultaneously using Multi-Kernel Blocks. Triplet Loss reduces classification errors in confusing transition segments by improving class separation. Sliding Window segmentation stabilizes model decisions. Clinically, this synergy allows the identification of transient abnormal episodes without sacrificing temporal consistency, explaining why combining the three components consistently improves performance.

These findings indicate that although each element contributes, multi-scale feature extraction and metric learning yield the greatest performance improvement, with temporal segmentation offering additional refinement. This shows that, despite the baseline model's high variability (±4-7%), the MK-TripNet classification is able to classify lung sounds with consistently low performance variation across various architecture configurations, with accuracy differences remaining within an acceptable range (<5%) and a small standard deviation (±1.64%). Although there were no significant differences between the advanced configurations, MK-TripNet statistically outperformed the baseline CNN with an accuracy improvement of up to 10.1%. These findings indicate that the combination of multi-scale convolution, metric learning, and temporal segmentation effectively achieves the highest accuracy of 89.1% while maintaining performance stability, which is crucial for practical clinical applications.

B. Hyperparameter Study

As shown in Table 3, this study evaluates the sensitivity of the proposed MK-tripNet model to various hyperparameter configurations, including embedding size, convolutional layer depth, pooling techniques, and dropout rate. To evaluate their impact on classification accuracy and generalization, one hyperparameter was changed at a time while the others remained constant, with an accuracy of 88.3 % and an F1 score of 0.88, the results show that an embedding size of 128 produces strong and stable performance, larger embedding sizes do not provide significant improvements, indicating that a moderate embedding size is sufficient to capture discriminative lung sound representations. Although performance gains diminish at greater convolutional depths, continuing to increase convolutional depth still improves performance, highlighting the importance of hierarchical feature extraction for modeling complex audio patterns. A moderate dropout rate of 0.3 produced the best generalization performance compared to higher dropout values (0.5), which caused underfitting, while max pooling performed slightly better than average pooling, indicating that retaining the most prominent features is more effective for lung sound classification.

Manuscript received 13 December 2025; Revised 20 February 2026; Accepted 15 March 2026; Available online 30 March 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1403

The robustness of the proposed design is confirmed by hyperparameter sensitivity analysis, which generally shows that MK-TripNet maintains stable performance across various configurations with accuracy fluctuations remaining within an acceptable range (<5%).

C. Quantitative Performance Comparison

The purpose of this study is to compare the proposed MK-TripNet model with baseline and alternative learning methods to determine whether there are significant quantitative performance differences in lung sound classification. To provide a comprehensive assessment of the model's effectiveness, the evaluation focused on standard classification metrics, including accuracy, precision, recall, and F1 score. Table 5 model's effectiveness summarizes the quantitative comparison results and shows the performance of MK-TripNet alongside several baseline and alternative models.

MK-TripNet consistently outperformed all comparison models on all evaluation metrics, as shown in Table 5. The proposed method's highest accuracy of 89.1%, precision of 84.5%, recall of 0.88, and F1 score of 0.89 demonstrate strong overall classification performance and balanced predictive power across all lung sound categories. The MFCC-CNN-BiGRU model achieved an accuracy of 83.5% and an F1 score of 0.82, highlighting the weakness of using cross-entropy loss and a single kernel convolution [7]. With an F1 score of 0.83 and an accuracy of 85.1%, the CNN-BiGRU-UMAP model showed a moderate improvement, indicating that dimension reduction contributed to feature compactness, while the performance improvement was still small compared to MK-TripNet [8].

In a similar context, a CNN with Triplet Loss achieved an F1 score of 0.83 and an accuracy 84.3%, indicating that metric learning improves discrimininative power but is insufficient when used without multi-scale feature extraction [14]. Without domain-specific architecture adaptation, general audio embeddings are less successful for lung sound classification, as seen in the VGGish-Triplet Loss model, which had the lowest performance (82.7% accuracy, 0.79% F1 score) [26]. Overall, MK-TripNet outperforms all baseline methods by an accuracy margin of between 3.2% and 6.4%, demonstrating the effectiveness of combining Sliding Window-based temporal segmentation, Triple Loss-based embedding learning, and Multi-Kernel convolution into a single framework. In addition, MK-TripNet successfully balanced recall (0.88) and precision (84.5%), demonstrating its ability to reduce both false positives and false negatives. This balanced performance indicates that the proposed architecture improves reliability across various lung sound categories while also improving classification accuracy. The stable performance improvement in Table 5 shows how robust MK-TripNet is compared to other learning techniques evaluated under the same experimental setting.

The model's ability to accurately detect 500 wheezing samples, as illustrated in Fig. 4., demonstrates that this system is highly viable for implementation in digital auscultation tools. This high level of sensitivity is crucial for reducing diagnostic uncertainty in clinical screenings. The quantitative performance of MK-TripNet has significant implications for computer-assisted auscultation systems and clinical lung sound analysis. To minimize false alarms and missed pathogens, automatic respiratory sound classification must not only achieve high accuracy but also strike the right balance between precision and recall [1], [2]. To facilitate clinical decision making and reduce diagnostic uncertainty during auscultation, the fine precision recall balance in MK-TripNet demonstrates better reliability in identifying abnormal lung sounds [3], [4].

Additionally, previous studies have shown that reliable temporal modeling and domain-specific feature extraction are important preconditions for useful lung sound analysis patterns to be present [5], [6]. In this regard, the superior quantitative performance of MK-TripNet demonstrates that the combination of discriminative embedding with multi-scale feature learning meets the clinical requirements for lung sound evaluation [7], [8]. These results indicate that MK-TripNet has significant potential for use in clinical care and real-world settings, where reliable lung sound classification and continuous respiratory monitoring depend on robustness, consistency, and reliability [2], [6].

V. Conclusion

This study proposed MK-TripNet, a novel deep learning architecture that integrates Multi-Kernel convolutional blocks, Triplet Loss-based embedding, and Sliding Window segmentation for real-time multi-class classification of lung auscultation sounds. The model achieved high performance with 89.1% accuracy, 0.89 F1-score, and 0.88 recall. These results consistently outperformed traditional and hybrid baselines, including MFCC-CNN-BiGRU, CNN-UMAP, and VGGish-Triplet Loss. In addition to accurate detection, the confusion matrix analysis confirmed that MK-TripNet is not only accurate but also robust in distinguishing between acoustically similar classes such as wheeze and crepitus. Future improvement should focus on cycle-aware attention mechanisms, applying self-supervised pretraining on larger datasets, extending the model to multi-label classification, and optimization for embedded or mobile deployment.

Manuscript received 13 December 2025; Revised 20 February 2026; Accepted 15 March 2026; Available online 30 March 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1403

Acknowledgment

The authors would like to express sincere gratitude to the Faculty of Engineering, Universitas Dian Nuswantoro, for the invaluable support and resources provided throughout this research. The facilities, academic environment, and encouragement from faculty members have significantly contributed to the completion of this work. This study would not have been possible without the institution's commitment to advancing research and innovation in medical electronics.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or non-profit sectors and was self-funded by the authors

Data Availability

No datasets were generated or analyzed during the current study.

Author Contribution

Widya Surya Erini conceived the main idea and concept of the study, collected the data, and performed the final editing of the manuscript. Gracia Putri Thomas contributed to the design and implementation of the machine learning model. Giulia Salzano Badia was responsible for programming, testing, and refining the machine learning model, as well as assisting in drafting the initial manuscript. Arief Rahadian provided medical and physiological expertise in respiratory acoustics, advised on the clinical interpretation of lung sound classifications, and validated the biomedical relevance of the proposed system. He also reviewed the methodology to ensure alignment with clinical diagnostic standards. Sofyan Budi Raharjo, M.D., Sp.P(K), FISR contributed clinical insights on pulmonary sounds, supervised the interpretation of auscultation-related findings, evaluated the model output from a respiratory medicine perspective, and ensured that the results were clinically meaningful and aligned with pulmonology practice guidelines. Sari Ayu Wulandari supervised the research process, provided guidance on developing the machine learning model, and validated the research findings. All authors have read, reviewed, and approved the final version of the manuscript for publication.

Declarations

Ethical Approval

This study did not involve human participants, animals, or personally identifiable data, and therefore, ethical approval was not required. The dataset used in this study was obtained from the publicly available Kaggle repository.

Consent for Publication Participants.

Not applicable. This study did not involve human participants.

Competing Interests

The authors declare no competing interests.

References

[1] Á. Troncoso, J. A. Ortega, R. Seepold, and N. M. Madrid, “Non-invasive devices for respiratory sound monitoring,” *Procedia Comput Sci*, vol. 192, pp. 3040–3048, 2021, doi: https://doi.org/10.1016/j.procs.2021.09.076.

[2] S. Ahamed Fayaz *et al.*, “Machine learning algorithms to predict treatment success for patients with pulmonary tuberculosis,” *PLoS One*, vol. 19, no. 10, p. e0309151, 2024, doi: https://doi.org/10.1371/journal.pone.0309151.

[3] A. Roy, B. Gyanchandani, A. Oza, and A. Singh, “TriSpectraKAN: a novel approach for COPD detection via lung sound analysis,” *Sci Rep*, vol. 15, no. 1, p. 6296, 2025, doi: https://doi.org/10.1038/s41598-024-82781-1.

[4] H. Duan *et al.*, “Global, Regional, and National Burden Trends in Chronic Obstructive Pulmonary Disease Attributable to Particulate Matter Pollution: 1990–2021 and Projections to 2036,” *Int J Chron Obstruct Pulmon Dis*, pp. 2671–2683, 2025, doi: https://doi.org/10.2147/COPD.S527263.

[5] D.-M. Huang, J. Huang, K. Qiao, N.-S. Zhong, H.-Z. Lu, and W.-J. Wang, “Deep learning-based lung sound analysis for intelligent stethoscope,” *Mil Med Res*, vol. 10, no. 1, p. 44, 2023, doi: https://doi.org/10.1186/s40779-023-00479-3.

[6] L. Hakkı and G. Serbes, “Detection of Wheeze Sounds in Respiratory Disorders: A Deep Learning Approach,” *International Advanced Researches and Engineering Journal*, vol. 8, no. 1, pp. 20–32, Apr. 2024, doi: 10.35860/iarej.1402462.

[7] M. Fraiwan, L. Fraiwan, B. Khassawneh, and A. Ibnian, “A dataset of lung sounds recorded from the chest wall using an electronic stethoscope,” *Data Brief*, vol. 35, p. 106913, 2021, doi: https://doi.org/10.1016/j.dib.2021.106913.

[8] S. Escobar-Pajoy and J. P. Ugarte, “Computerized analysis of pulmonary sounds using uniform manifold projection,” *Chaos Solitons Fractals*, vol. 166, p. 112930, 2023, doi: https://doi.org/10.1016/j.chaos.2022.112930.

[9] A. H. Sfayyih, N. Sulaiman, and A. H. Sabry, “A review on lung disease recognition by acoustic signal analysis with deep learning networks,” *J*

Manuscript received 13 December 2025; Revised 20 February 2026; Accepted 15 March 2026; Available online 30 March 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1403

Big Data, vol. 10, no. 1, p. 101, 2023, doi: https://doi.org/10.1186/s40537-023-00762-z.

[10] R. Zulfiqar, F. Majeed, R. Irfan, H. T. Rauf, E. Benkhelifa, and A. N. Belkacem, "Abnormal respiratory sounds classification using deep CNN through artificial noise addition," *Front Med (Lausanne)*, vol. 8, p. 714811, 2021, doi: https://doi.org/10.3389/fmed.2021.714811.

[11] G. Petmezas *et al.*, "Automated lung sound classification using a hybrid CNN-LSTM network and focal loss function," *Sensors*, vol. 22, no. 3, p. 1232, 2022, doi: https://doi.org/10.3390/s22031232.

[12] F.-S. Hsu *et al.*, "A dual-purpose deep learning model for auscultated lung and tracheal sound analysis based on mixed set training," *Biomed Signal Process Control*, vol. 86, p. 105222, 2023, doi: https://doi.org/10.1016/j.bspc.2023.105222.

[13] R. Khan, S. U. Khan, U. Saeed, and I.-S. Koo, "Auscultation-based pulmonary disease detection through parallel transformation and deep learning," *Bioengineering*, vol. 11, no. 6, p. 586, 2024, doi: https://doi.org/10.3390/bioengineering11060586.

[14] P. Duangmanee *et al.*, "Triplet Multi-Kernel CNN for Detection of Pulmonary Diseases From Lung Sound Signals," *IEEE Access*, 2025, doi: http://dx.doi.org/10.1109/ACCESS.2025.3552108.

[15] S. A. Shehab, K. K. Mohammed, A. Darwish, and A. E. Hassanien, "Deep learning and feature fusion-based lung sound recognition model to diagnoses the respiratory diseases," *Soft comput*, vol. 28, no. 19, pp. 11667–11683, 2024, doi: https://doi.org/10.1007/s00500-024-09866-x.

[16] T. Wanasinghe, S. Bandara, S. Madusanka, D. Meedeniya, M. Bandara, and I. D. L. T. Díez, "Lung sound classification with multi-feature integration utilizing lightweight CNN model," *IEEE Access*, vol. 12, pp. 21262–21276, 2024, doi: https://doi.org/10.1109/ACCESS.2024.3361943.

[17] T. Nguyen and F. Pernkopf, "Crackle detection in lung sounds using transfer learning and multi-input convolutional neural networks," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE, 2021, pp. 80–83. doi: https://doi.org/10.1109/EMBC46164.2021.9630577.

[18] Y. Kim, K. B. Kim, A. Y. Leem, K. Kim, and S. H. Lee, "Enhanced Respiratory Sound Classification Using Deep Learning and Multi-Channel Auscultation," *J Clin Med*, vol. 14, no. 15, p. 5437, 2025, doi: https://doi.org/10.3390/jcm14155437.

[19] J. Li *et al.*, "LungAttn: advanced lung sound classification using attention mechanism with dual TQWT and triple STFT spectrogram," *Physiol Meas*, vol. 42, no. 10, p. 105006, 2021, doi: https://doi.org/10.3390/arm93050032.

[20] S. B. Shuvo, S. N. Ali, S. I. Swapnil, T. Hasan, and M. I. H. Bhuiyan, "A lightweight CNN model for detecting respiratory diseases from lung auscultation sounds using EMD-CWT-based hybrid scalogram," *IEEE J Biomed Health Inform*, vol. 25, no. 7, pp. 2595–2603, 2020, doi: https://doi.org/10.1109/jbhi.2020.3048006.

[21] F.-S. Hsu, S.-R. Huang, C.-W. Huang, C.-C. Chen, Y.-R. Cheng, and F. Lai, "Multi-path Convolutional Neural Networks Efficiently Improve Feature Extraction in Continuous Adventitious Lung Sound Detection," *arXiv preprint arXiv:2107.04226*, 2021, doi: https://doi.org/10.48550/arXiv.2107.04226.

[22] N. Fraihi, O. Karrakchou, and M. Ghogho, "Improving Deep Learning-based Respiratory Sound Analysis with Frequency Selection and Attention Mechanism," *arXiv preprint arXiv:2507.20052*, 2025, doi: https://doi.org/10.48550/arXiv.2507.20052.

[23] L. Pham, D. Ngo, A. Schindler, and R. King, "A Deep Neural Network with Triplet Loss for Detecting Anomaly of Respiratory Sounds" in Proc. *DAGA 2021 (47th Annual Meeting for Acoustics)*, Vienna, Austria, 2021, ISBN: 978-3-939296-18-8

[24] S. Abhishek, A. J. Ananthapadmanabhan, T. Anjali, S. Reyma, A. Perathur, and R. B. Bentov, "Multimodal Integration of an Enhanced Novel Pulmonary Auscultation Real-Time Diagnostic System," *IEEE MultiMedia*, vol. 31, no. 3, pp. 18–43, 2024, doi: https://doi.org/10.1109/MMUL.2024.3422022.

[25] A. Sadeghzadeh and M. B. Islam, "Triplet loss-based convolutional neural network for static sign language recognition," in *2022 Innovations in Intelligent Systems and Applications Conference (ASYU)*, IEEE, 2022, pp. 1–6. doi: https://doi.org/10.1109/ASYU56188.2022.9925490.

[26] Y. Kim *et al.*, "Respiratory sound classification for crackles, wheezes, and rhonchi in the clinical field using deep learning," *Sci Rep*, vol. 11, no. 1, p. 17186, 2021, doi: https://doi.org/10.1038/s41598-021-96724-7.

[27] E. Messner *et al.*, "Multi-channel lung sound classification with convolutional recurrent neural networks," *Comput Biol Med*, vol. 122, p. 103831, 2020, doi:

Manuscript received 13 December 2025; Revised 20 February 2026; Accepted 15 March 2026; Available online 30 March 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1403

https://doi.org/10.1016/j.compbiomed.2020.103831.
[28] S. KV, D. Koppad, P. Kumar, N. A. Kantikar, and S. Ramesh, "Multi-Task Learning for Lung sound & Lung disease classification," *arXiv preprint arXiv:2404.03908*, 2024, doi: https://doi.org/10.48550/arXiv.2404.03908.
[29] F. Wang, X. Yuan, Y. Liu, and C.-T. Lam, "LungNeXt: A novel lightweight network utilizing enhanced mel-spectrogram for lung sound classification," *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 8, p. 102200, Oct. 2024, doi: 10.1016/j.jksuci.2024.102200.
[30] J. Park, C. Jeong, Y. Choi, H. Hong, and Y. Jo, "Lung Sound Classification Model for On-Device AI," *Applied Sciences*, vol. 15, no. 17, p. 9361, 2025, doi: https://doi.org/10.3390/app15179361.
[31] Y. Zhang, Q. Huang, W. Sun, F. Chen, D. Lin, and F. Chen, "Research on lung sound classification model based on dual-channel CNN-LSTM algorithm," *Biomed Signal Process Control*, vol. 94, p. 106257, 2024, doi: https://doi.org/10.1016/j.bspc.2024.106257.
[32] Z. Wang and Z. Sun, "Performance evaluation of lung sounds classification using deep learning under variable parameters," *EURASIP J Adv Signal Process*, vol. 2024, no. 1, p. 51, 2024, doi: https://doi.org/10.1186/s13634-024-01148-w.
[33] A. Mallol-Ragolta, M. Milling, and B. Schuller, "Multi-triplet loss-based models for categorical depression recognition from speech," in *Proceedings of the 7th Iberian Speech and Language Technologies Conference*, 2024, pp. 31–35. doi: 10.21437/IberSPEECH.2024-7.
[34] A. Shevchyk, R. Hu, K. Thandiackal, M. Heizmann, and T. Brunschwiler, "Privacy preserving synthetic respiratory sounds for class incremental learning," *Smart Health*, vol. 23, p. 100232, 2022, doi: https://doi.org/10.1016/j.smhl.2021.100232.
[35] B. Shen, M. Zhang, L. Yao, and Z. Song, "Novel triplet loss-based domain generalization network for bearing fault diagnosis with unseen load condition," *Processes*, vol. 12, no. 5, p. 882, 2024, doi: https://doi.org/10.3390/pr12050882.
[36] T. R. Ornob, G. Roy, and E. Hassan, "CovidExpert: A Triplet Siamese Neural Network framework for the detection of COVID-19," *Inform Med Unlocked*, vol. 37, p. 101156, 2023, doi: https://doi.org/10.1016/j.imu.2022.101156.
[37] Y. Kono *et al.*, "Breath measurement method for synchronized reproduction of biological tones in an augmented reality auscultation training system," *Sensors*, vol. 24, no. 5, p. 1626, 2024, doi: https://doi.org/10.3390/s24051626.
[38] Y. Torabi, S. Shirani, and J. P. Reilly, "Descriptor: Heart and Lung Sounds Dataset Recorded from a Clinical Manikin using Digital Stethoscope (HLS-CMDS)," *IEEE Data Descriptions*, 2025, doi:https://doi.org/10.1109/IEEEDATA.2025.3566012.
[39] A. H. Sabry, O. I. Dallal Bashi, N. H. Nik Ali, and Y. Mahmood Al Kubaisi, "Lung disease recognition methods using audio-based analysis with machine learning," Feb. 29, 2024, *Elsevier Ltd*. doi: https://doi.org/10.1016/j.heliyon.2024.e26218.
[40] P. Kapetanidis *et al.*, "Respiratory Diseases Diagnosis Using Audio Analysis and Artificial Intelligence: A Systematic Review," *Sensors*, vol. 24, no. 4, Feb. 2024, doi: https://doi.org/10.3390/s24041173.

Author Biography

Widya Surya Erini is an undergraduate student in the Department of Biomedical Engineering, Universitas Dian Nuswantoro. She is currently completing her thesis on the development of a digital stethoscope using the MK-TripNet method. She gained research experience during an internship at the Indonesian Medical Education and Research Institute (IMERI) at Universitas Indonesia in 2024. She has served as the President of the Biomedical Engineering Student Association in Universitas Dian Nuswantoro and as the team leader whose proposal successfully received funding from the Ministry of Education and EPICS IEEE 2025. In 2025, she was also selected as a finalist in the Respiratory Care Championship (RespiChamp). Her research interests include biomedical engineering, biomedical signal processing, machine learning applications, and health technology innovation.

Gracia Putri Thomas is an undergraduate student in the Department of Biomedical Engineering, Universitas Dian Nuswantoro. She is currently completing her thesis on the development of a digital stethoscope using the MK-TripNet method. In 2024, she gained research experience through an internship at the Indonesian Medical Education and Research Institute (IMERI), Universitas Indonesia. She has received funding from the Indonesia Student Creativity Program (PKM) 2025 and EPICS IEEE 2025, and has actively taken leadership roles in student organizations. In

Manuscript received 13 December 2025; Revised 20 February 2026; Accepted 15 March 2026; Available online 30 March 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1403

2025, she was also selected as a finalist in the Respiratory Care Championship (RespiChamp). Her research interests include biomedical engineering, biomedical signal processing, machine learning applications, and health technology innovation.

Giulia Salzano Badia is an undergraduate student in the Department of Biomedical Engineering, Universitas Dian Nuswantoro. She is currently working on her thesis, focusing on developing a digital stethoscope using the MK-TripNet method. In 2024, she gained research experience through an internship at the Indonesian Medical Education and Research Institute (IMERI), Universitas Indonesia. She has received funding from the Student Creativity Program (PKM) in 2024 and 2025, as well as from EPICS IEEE in 2025. In the same year, she was also selected as a finalist in the Respiratory Care Championship (RespiChamp). Her research interests include biomedical engineering, medical device design, and healthcare technological innovation.

Arief Rahadian is a medical lecturer in the Department of Medicine at Universitas Dian Nuswantoro. He earned his Ph.D. in Cardiovascular Medicine from Tokushima University, Japan, where he also served as a teaching and research assistant. His research focuses on endothelial dysfunction, atherosclerosis, and molecular mechanisms of cardiovascular diseases. He has published in journals and presented at international conferences, including those of the Japanese Circulatory Society and the European Society of Cardiology. He has received several research and academic awards, including the Fuji Otsuka Foundation Scholarship and the Young Researcher Award at the Tokushima Bioscience Retreat. His research interests include cardiovascular biology, molecular medicine, and translational biomedical research.

Sofyan Budi Raharjo, M.D., Sp.P(K), FISR, is a pulmonologist and respiratory medicine consultant at Dr. Kariadi General Hospital, Semarang, Indonesia. He earned his medical degree from Sultan Agung Islamic University in 2004 and completed his specialization in Pulmonology and Respiratory Medicine at Sebelas Maret University in 2010, and obtained his consultancy in Pulmonary Intervention and Critical Respiratory Care in 2022. He has participated in various national and international training programs, including pulmonary intervention workshops under the European Respiratory Society (ERS). He currently serves as the Chairman of the Indonesia Society of Respirology (PDPI) Central Java Branch and Head of the Medical Division of the Indonesia Asthma Foundation, Semarang Branch. His research and clinical interests focus on pulmonary intervention, intensive respiratory care, and thoracic imaging.

Sari Ayu Wulandari (Student Member, IEEE) obtained a master's degree in Electrical Engineering from Gadjah Mada University in 2011. From 2021 until the present, she has been pursuing a doctoral degree at Sepuluh Nopember Institute of Technology. She is currently a lecturer in Biomedical Engineering at Universitas Dian Nuswantoro and a member of the Cemti Laboratory (Center for Medical Technology Innovation). Her research interests lie in biomedical engineering, particularly in signal and image processing, with a focus on deep learning systems.

Manuscript received 13 December 2025; Revised 20 February 2026; Accepted 15 March 2026; Available online 30 March 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1403