

ภาคผนวก ก

บทความวิชาการที่ได้รับการตีพิมพ์เผยแพร่ในระหว่างการศึกษา

รายชื่อบทความวิชาการที่ได้รับการตีพิมพ์เผยแพร่ระหว่างการศึกษา

- Orkweha, K., Phapatanaburi, K., Pathonsuwan, W., Jumphoo, T., Rattanasak, A., Anchuen, P., Pinthurat, W., Uthansakul, M., and Uthansakul, P. (2024). A framework for detecting pulmonary diseases from lung sound signals using a hybrid multi-task autoencoder-SVM model. *Symmetry*, 16(11), 1413. <https://doi.org/10.3390/sym16111413>
- Duangmanee, P., Phapatanaburi, K., Pathonsuwan, W., Jumphoo, T., Rattanasak, A., Orkweha, K., Anchuen, P., Uthansakul, M., and Uthansakul, P. (2025). Triplet multi-kernel CNN for detection of pulmonary diseases from lung sound signals. *IEEE Access*, 13, 50137–50151. <https://doi.org/10.1109/ACCESS.2025.3552108>

Article

A Framework for Detecting Pulmonary Diseases from Lung Sound Signals Using a Hybrid Multi-Task Autoencoder-SVM Model

Khwanjit Orkweha ¹, Khomdet Phapatanaburi ^{2,*}, Wongsathon Pathonsuwan ³, Talit Jumphoo ³, Atcharawan Rattanasak ¹, Patikorn Anchuen ⁴, Watcharakorn Pinthurat ⁵, Monthippa Uthansakul ¹ and Peerapong Uthansakul ^{1,*}

¹ School of Telecommunication Engineering, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand; d6500597@g.sut.ac.th (K.O.); d6500658@g.sut.ac.th (A.R.); mtp@sut.ac.th (M.U.)

² Department of Telecommunication Engineering, Faculty of Engineering and Technology, Rajamangala University of Technology Isan (RMUTI), Nakhon Ratchasima 30000, Thailand

³ Institute of Research and Development, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand; pathonsuwan.wst@sut.ac.th (W.P.); jumphoo@sut.ac.th (T.J.)

⁴ Navaminda Kasatriyadhiraj Royal Air Force Academy, Saraburi 18180, Thailand; patikorn_an@taf.mi.th

⁵ Department of Electrical Engineering, Rajamangala University of Technology Tawan-Ok, Chanthaburi 22210, Thailand; watcharakorn_pi@rmuto.ac.th

* Correspondence: khomdet.ph@rmuti.ac.th (K.P.); uthansakul@sut.ac.th (P.U.)

Abstract: Research focuses on the efficacy of Multi-Task Autoencoder (MTAE) models in signal classification due to their ability to handle many tasks while improving feature extraction. However, researchers have not thoroughly investigated the study of lung sounds (LSs) for pulmonary disease detection. This paper introduces a new framework that utilizes an MTAE model to detect lung diseases based on LS signals. The model integrates an autoencoder and a supervised classifier, simultaneously optimizing both classification accuracy and signal reconstruction. Furthermore, we propose a hybrid approach that combines an MTAE and a Support Vector Machine (MTAE-SVM) to enhance performance. We evaluated our model using LS signals from a publicly available database from King Abdullah University Hospital. The model attained an accuracy of 89.47% for four classes (normal, pneumonia, asthma, and chronic obstructive pulmonary disease) and 90.22% for three classes (normal, pneumonia, and asthma cases). Using the MTAE-SVM, the accuracy was further improved to 91.49% for four classes and 93.08% for three classes, respectively. The results indicate that the MTAE and MTAE-SVM have a considerable potential for detecting pulmonary diseases from lung sound signals. This could aid in the creation of more user-friendly and effective diagnostic tools.

Keywords: pulmonary disease detection; multi-task autoencoder; MFCC; audio diagnostics; health informatics



Citation: Orkweha, K.; Phapatanaburi, K.; Pathonsuwan, W.; Jumphoo, T.; Rattanasak, A.; Anchuen, P.; Pinthurat, W.; Uthansakul, M.; Uthansakul, P. A Framework for Detecting Pulmonary Diseases from Lung Sound Signals Using a Hybrid Multi-Task Autoencoder-SVM Model. *Symmetry* **2024**, *16*, 1413. <https://doi.org/10.3390/sym16111413>

Academic Editor: Tomohiro Inagaki

Received: 12 September 2024

Revised: 17 October 2024

Accepted: 22 October 2024

Published: 23 October 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Breathing-generated lung sounds (LSs) carry valuable health information [1] crucial for diagnosing lung illnesses [2]. Given the highly transmissible and potentially severe nature of many diseases, prompt and accurate identification is paramount. However, this task is challenging as LSs often occur at frequencies beyond human hearing, increasing the risk of missing important diagnostic cues. Furthermore, the assessment of LSs, known as auscultation, by healthcare practitioners is inherently subjective and may lead to diagnostic errors [3]. To tackle these challenges [4], there is a growing trend towards adopting computer-aided diagnosis methods. These technologies offer a more objective and reliable approach to diagnosing respiratory disorders, enhancing the accuracy and efficiency of lung disease diagnostics. In this paper, we focus on a framework for detecting pulmonary diseases (PDs) from LS signals, leveraging pattern recognition techniques in the field

of computer science. This approach falls within the broader domain of symmetry in computational models.

In recent years, significant advancements have been made in detecting PDs using LS signals and conventional machine learning techniques. The typical structure of these LS detection systems comprises two main phases: initial feature extraction and subsequent classification. During feature extraction, these systems meticulously analyze breathing sounds, employing various classification models. Previous studies have demonstrated a preference for wavelet-based features [5–7], Mel Frequency Cepstral Coefficients (MFCCs) [8–11], Linear Predictive Cepstral Coefficients (LPCCs) [12], bispectrum [8], and Empirical Wavelet Transform with Fixed Boundary Points (EWT-FBP) [13]. Particularly, MFCCs have gained widespread adoption due to their exceptional ability to capture crucial spectral characteristics of LSs, essential for distinguishing between different pulmonary conditions. They provide a compact yet comprehensive representation of audio signals, facilitating efficient analysis. Their small size and robustness in noisy environments make them ideal for clinical settings where LS recordings may contain background noise. In the classification phase, various models including Support Vector Machines (SVMs) [5,12,14], Gaussian Mixture Models (GMMs) [15,16], k-Nearest Neighbors (KNN) [17], Linear Discriminant Analysis (LDA) [17], Decision Trees (DTs) [18], Multilayer Perceptron (MLP) [13], Random Forest (RF) [13,19], Extreme Gradient Boosting (XGBoost) [13], and Light Gradient Boosting Machine (LGBM) [13,16,17] have been utilized. The choice of classifier significantly impacts the system's effectiveness in diagnosing PDs from LSs, underscoring the importance of proficient feature extraction design. Therefore, the development of these systems necessitates expertise in feature extraction methodologies to ensure accurate and reliable detection of pulmonary conditions.

In addition to the classifiers previously discussed, the utilization of Deep Neural Network (DNN) methods in detecting PDs has gained traction due to their growing effectiveness [20,21]. A noteworthy development in this domain is the application of Convolutional Neural Networks (CNNs) for LS analysis, elucidated in [22]. This approach commences with the extraction of MFCC features, which are then inputted into a two-layer CNN (CNN-2L) for classification. The CNN-2L has demonstrated superior performance compared with traditional SVM-based techniques, particularly in distinguishing normal LSs from those indicative of asthma or pneumonia. This success underscores the potential of combining CNNs with MFCCs for more precise classification. The CNN-2L serves as a benchmark model in this study for PDs detection from LS signals. Furthermore, there has been notable progress in merging DNNs with Hidden Markov Models (DNN-HMMs), as detailed in [23]. This fusion harnesses the predictive capabilities of DNNs and the proficiency of HMMs in recognizing temporal patterns, significantly enhancing classification performance. Another significant advancement is the integration of the VGGish network with the bidirectional gated recurrent unit neural network (BiGRU), as proposed in [24]. This hybrid model utilizes Mel-spectrogram features, capitalizing on the VGGish network's advanced feature extraction and the BiGRU's capability to process sequential data. The VGGish-BiGRU model exemplifies the benefits of amalgamating these advanced neural network architectures, leading to enhanced accuracy and reliability in LS analysis for disease identification, thereby paving the way for future advancements.

Deep learning algorithms have demonstrated remarkable potential in accurately identifying diseases from LS signals. However, these systems often focus on single tasks, potentially overlooking critical information. To address this issue, researchers have developed Multi-Task Autoencoders (MTAEs), aiming to simultaneously optimize both the reconstruction fidelity of the input signal and the accuracy of disease classification. This dual-focus approach ensures not only precise disease detection but also the preservation of detailed information from the original LSs, essential for complex medical diagnostics. The versatility of MTAEs is further evidenced by their successful deployment in various domains, demonstrating their robustness and adaptability in handling intricate data-driven challenges. A notable example is the Event-Related Potential Encoder Network

(ERPENet) [25], which integrates CNNs with Long Short-Term Memory (LSTM) within an autoencoder architecture to efficiently compress Electroencephalography (EEG) signals and extract critical P300 features, essential for tasks like the P300 speller. Similarly, a novel end-to-end multi-task learning method, referred to as MIN2Net, was introduced in [26], which represents an advanced MTAE model for independent motor imagery EEG classification, outperforming standard benchmarks in brain-machine interfaces on the SMR-BCI and OpenBMI datasets. Another significant advancement is the multi-task convolutional autoencoder-based neural network for brain tumor classification, as reported in [27]. Applied to the Cheng dataset, this method achieves remarkable accuracy, categorizing brain tumors into meningioma, glioma, and pituitary with a 99.30% success rate. These instances collectively underscore the sophisticated capabilities and adaptability of MTAEs in managing complex, data-centric tasks. However, their application in PD detection using LS signals remains uncharted territory, offering opportunities for significant advancements in medical diagnostics.

In this paper, we introduce an innovative framework utilizing an MTAE, specifically tailored for comprehensive analysis of LS signals. This novel framework integrates an autoencoder and a supervised classifier in a synergistic combination. The autoencoder, comprising an encoder and a decoder, is adept at creating a latent vector that encapsulates the fundamental characteristics of LSs. Simultaneously, the supervised classifier interprets this latent vector to extract clinically significant information. This model meticulously designs its training process to simultaneously optimize two crucial factors: the accuracy of disease classification and the fidelity of signal reconstruction. This dual-focus approach is pivotal for enhancing the model's capability to accurately detect various lung diseases while preserving the integrity of the original LS signals. We were inspired by [28,29], which combined a CNN/Raw Speech-Based Multi-Scale Convolutional Neural Network (RS-MSCConvNet) for trainable feature representation with an SVM as the classifier. This model was more accurate than CNN/RS MSCConvNet alone, which was mostly due to its ability to use both the CNN's feature extraction strengths and the SVM's classification strengths. In particular, the distance between the CNN/RS-MSCConvNet features that were automatically extracted and the trained SVM's hyperplane made it more specific, which improved its ability to detect. In this paper, we propose a hybrid MTAE-SVM model that replaces the final fully connected layers of the MTAE after full training, resulting in further improvements in classification accuracy.

In this paper, we introduce an innovative framework utilizing an MTAE, specifically tailored for comprehensive analysis of LS signals. This novel framework integrates an autoencoder and a supervised classifier in a synergistic combination. We chose the MTAE model for its distinctive capacity to simultaneously improve classification accuracy and signal reconstruction quality, which is essential for the effective analysis of complex patterns in LS data. Our model addresses the shortcomings of current methods that concentrate only on classification, which could miss essential diagnostic information comprised in the LS signals. The autoencoder, comprising an encoder and a decoder, is adept at creating a latent vector that encapsulates the fundamental characteristics of LSs. Simultaneously, the supervised classifier interprets this latent vector to extract clinically significant information. This model meticulously designs its training process to simultaneously optimize two crucial factors: the accuracy of disease classification and the fidelity of signal reconstruction. This dual-focus approach is pivotal for enhancing the model's capability to accurately detect various lung diseases while preserving the integrity of the original LS signals. Our experimental results demonstrate that the proposed MTAE model achieves greater accuracy in PD detection from LS waves, notably outperforming current approaches, including the benchmark CNN-2L model. This increase in performance demonstrates the efficacy of our method and its potential to advance medical diagnostics.

The salient contributions of this paper can be summarized as:

1. This study introduces the inaugural design of a novel framework utilizing MTAE models tailored specifically for the detection of PDs from LS signals. To the best of

our knowledge, this is the first attempt to introduce a framework that integrates both regression and classification methods for the detection of PDs using LS signals.

2. We investigate our model on the dataset provided by [30] from King Abdullah University Hospital (KAUH). Our framework demonstrates remarkable accuracy, achieving 89.47% reliability in distinguishing between normal, pneumonia, asthma, and chronic obstructive pulmonary disease (COPD) using fivefold cross-validation on the KAUH dataset. Additionally, it attains a 90.22% accuracy in classifying normal, pneumonia, and asthma cases.
3. The MTAE model is further refined by incorporating an SVM method, replacing the final fully connected layer to classify the automatically extracted features from the fully trained MTAE. This modified version, referred to as the hybrid MTAE-SVM model architecture, delivers enhanced accuracy compared with the MTAE classifier alone.

The subsequent sections of the paper are structured as follows. Related work is presented and discussed in Section 2. The proposed framework is explained in Section 3, while the experimental setup and the obtained results are deliberated in Sections 4 and 5, respectively. Finally, Section 6 concludes the paper and offers possible directions for future investigation.

2. Related Works

This section offers a concise overview of commonly employed CNN methodologies for detecting PDs using LS signals, which will serve as our baseline for comparison with our proposed approach. The emphasis lies on utilizing MFCCs for feature extraction, followed by CNNs for classification.

2.1. MFCC

The MFCC is pivotal in audio signal processing, especially in capturing the short-term power spectrum of a sound. Obtaining the MFCC involves a sequence of successive steps. Initially, the process utilizes the Fourier Transform to generate the power spectrum. Then, the spectra are mapped onto the mel scale, which reflects the perceived frequency of a tone. This is followed by logarithmic scaling and ultimately culminates in the application of the inverse Fourier Transform. The mathematical formulation of the MFCC is presented as,

$$MFCC(i) = \sum_{k=0}^{N-1} \log(P(k)) \cdot \cos\left((2k+1) \frac{\pi i}{2N}\right), \quad (1)$$

for $i = 1, 2, \dots, M$,

where $P(k)$ denotes the power spectrum, N represents the count of mel filters, and M indicates the number of cepstral coefficients.

Following the approach outlined in [31], we resized the initial MFCC images to a more practical 28×28 pixel grayscale format. This resizing was intentionally performed to reduce computational burden while preserving data integrity.

2.2. CNN-Based Classifier

CNNs play a pivotal role in classifying PDs using LS signals. Following feature extraction using the MFCC, the processed data are fed into a CNN, renowned for its exceptional proficiency in detecting patterns and conducting classification tasks. Specifically designed to scrutinize the precise features of the sound signals represented in MFCC images, the CNN architecture comprises convolutional layers, pooling layers, and fully connected (FC) layers arranged hierarchically. Each layer of the CNN contributes to the detection process by extracting and enhancing features from the MFCC data, ultimately leading to robust and precise classification of LSs.

The CNN methodology commences with convolutional layers, which utilize filters on the input MFCC images to produce feature maps, highlighting significant characteristics in the sound signals. This approach is pivotal for discerning the intricate patterns

in LSs indicative of various lung diseases. Subsequently, max-pooling layers reduce the dimensionality of the feature maps, thereby enhancing the computational efficiency of the network and mitigating the risk of overfitting. The final layers of the CNN are densely connected, enabling advanced reasoning processes based on the extracted features. These layers amalgamate information from the feature maps to ultimately classify the LSs, distinguishing between healthy and diseased conditions. The CNN architecture configuration can be observed in Figure 1.

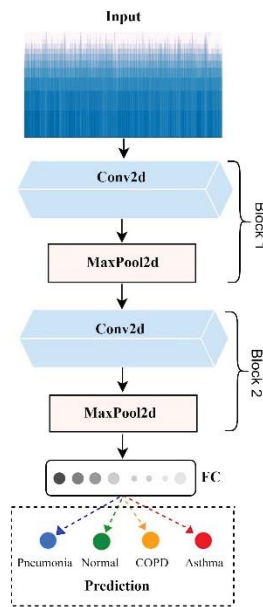


Figure 1. The configuration of CNN architecture.

3. Proposed Framework

This section presents our innovative approach for detecting PDs from LS signals using the proposed MTAE model and the MTAE-SVM hybrid architecture. The methodology is structured into three key phases: data preparation and feature extraction, the design and training of the MTAE model, and the integration of the MTAE-SVM hybrid model architecture.

3.1. Data Preparation and Feature Extraction

The dataset comprises audio recordings demonstrating various types of PDs, with each recording indexed by its associated file path and disease label. The preprocessing step involves applying a Butterworth bandpass filter to the audio data. The transfer function of the filter is given by,

$$H(f) = \frac{1}{\sqrt{1 + \left(\frac{f}{f_c}\right)^{2n}}}, \quad (2)$$

where n represents the filter order (set to 12), f is the frequency, and f_c is the cut-off frequencies, ranging between 120 Hz and 1800 Hz. Following this, MFCCs, as elaborated in the preceding section, are extracted from these filtered signals.

3.2. MTAE Model Architecture

This section presents a novel MTAE model meticulously crafted as a sophisticated neural network consisting of two primary components: an autoencoder for precise signal reconstruction and a classifier for accurate disease identification. These elements work in tandem to analyze LS signals, offering vital insights for advanced medical diagnostics. The structure of the proposed MTAE framework is depicted in Figure 2.

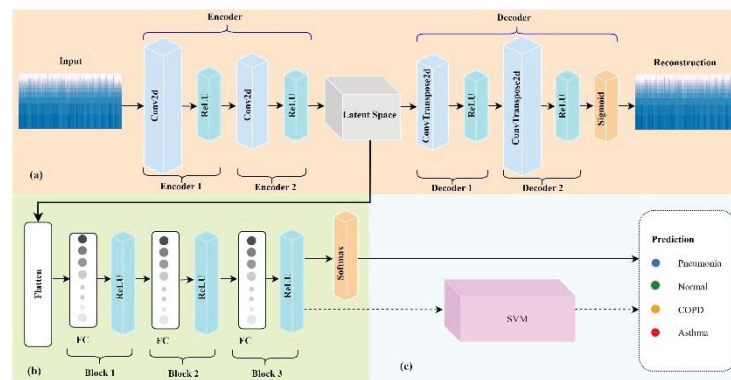


Figure 2. The architecture of our proposed framework, illustrating: (a) the autoencoder for unsupervised feature learning, (b) the MTAE-based classifier for multi-task representation learning, and (c) the MTAE-SVM hybrid classifier, which integrates MTAE features with an SVM for final classification.

3.2.1. Autoencoder

The autoencoder, an integral component of our strategic framework, has been meticulously crafted to transform intricate LS data into a condensed, lower-dimensional latent vector, facilitating swift feature extraction. This transformation process is essential for capturing inherent patterns within the data, crucial for precise disease detection. The autoencoder systematically restores the data to their original high dimensionality, thereby preserving the quality of the signals.

Encoder

The encoder plays a pivotal role in constraining the input data. It adeptly captures and encodes the most crucial characteristics into a hidden space representation, a critical step for effective pattern recognition. This is achieved through a series of convolutional layers, each activated by a Rectified Linear Unit (ReLU) function. The ReLU function, denoted as $\text{ReLU}(x) = \max(0, x)$, introduces non-linearity, enabling the model to better analyze intricate and non-linear relationships observed in LS data. By exclusively preserving positive values, this function significantly enhances the model's learning process and addresses challenges like the vanishing gradient problem, commonly encountered in DNN training.

Decoder

The decoder functions as the counterpart to the encoder, aiming to accurately reconstruct the original input from its compressed latent representation. This task is achieved by employing transposed convolutional layers, each accompanied by ReLU activations. These layers progressively enhance the resolution of the compressed, data-dense latent vector to align with the original input, maintaining a high number of dimensions. The symmetrical arrangement of the encoder and decoder ensures the retention of vital data

characteristics, enabling the model to obtain a concise yet comprehensive representation of the input. This capability proves particularly advantageous for unsupervised learning tasks, such as anomaly identification and noise reduction in LS analysis, where grasping the underlying structure of the data is paramount.

3.2.2. MTAE-Based Classifier

The classifier is intricately linked to the latent vector of the autoencoder and has been meticulously crafted for the specific task of disease classification. Comprising multiple linear layers, the system is adeptly structured to interpret the encoder's output, thus transforming it into precise disease detection results. The architecture of the classifier is meticulously designed to enhance its capability to discern subtle variations in LS patterns, which serve as indicators of different PDs. The output of the classifier plays a crucial role in computing the cross-entropy loss, a fundamental metric for training and refining the model. With its emphasis on precision and accuracy in disease detection, the classifier emerges as an indispensable component in the diagnostic process.

3.2.3. Training Procedure for MTAE

Throughout the training phase, our model is meticulously engineered to achieve two primary objectives: precise disease classification and effective signal reconstruction. Through the utilization of a dual-target approach, the model's ability to generalize to unfamiliar data and uphold accuracy in disease categorization is enhanced.

The focus of the training revolves around optimizing a composite loss function, which is defined as,

$$L = L_{recon} + L_{class}, \quad (3)$$

where L_{recon} represents the reconstruction loss, quantified by the Mean Squared Error (MSE), and L_{class} is the classification loss, determined using the cross-entropy (CE) loss. These components are crucial in steering the model's learning process:

- **MSE for reconstruction:** this loss function evaluates the average squared discrepancies between the reconstructed signals and the original inputs. It is formulated as,

$$L_{recon} = \frac{1}{S} \sum_{i=1}^S (Y_i - \hat{Y}_i)^2, \quad (4)$$

where S is the total number of samples, Y_i is the actual signal value, and $\hat{Y}_i$ the reconstructed version. The MSE is vital for ensuring the model accurately reproduces the input signals, a key aspect for effective diagnosis.

- **CE for classification:** this loss assesses the model's accuracy in disease classification and is formulated as,

$$L_{class} = - \sum_{c=1}^C y_{oc} \log(p_{oc}), \quad (5)$$

where C is the number of disease categories, y is a binary indicator (1 if the class label c is correct for observation o , 0 otherwise), and p is the predicted probability of observation o belonging to class c . This loss refines the model's capacity to distinguish between diseases.

By merging these two loss functions, the model is effectively guided towards accurate disease categorization while simultaneously proficiently reconstructing input signals. Implementing this well-balanced strategy is pivotal in crafting a robust model highly suitable for practical diagnostic applications. This joint optimization implies that the common latent representation includes features that are both distinctive for classification and informative for reconstruction, as suggested in [32,33]. Consequently, the model augments its capacity to identify small deviations in LS signals, thereby enhancing classification accuracy while

preserving high-quality signal reconstruction, which is crucial for the development of dependable diagnostic tools.

3.3. MTAE-SVM Model Architecture

The SVM-based approach has proven effective in both classification and regression tasks [34], making it widely applicable across various domains [35–37], including image classification [38] and audio classification [39]. In addition, the SVM-based method is one popular method for detecting pulmonary diseases due to its robust handling of complex, high-dimensional feature spaces, which are common in medical signal processing. Previous studies [28,29] motivated us by demonstrating the efficacy of integrating a CNN or an RS-MSCConvNet for feature extraction alongside an SVM classifier. The hybrid models demonstrated superior accuracy compared with utilizing a CNN or RS-MSCConvNet independently, primarily due to their ability to utilize the advantages of the CNN's feature extraction alongside the SVM's classification strengths. The features extracted automatically by the CNN/RS-MSCConvNet, when fed into the SVM, leveraged the SVM's ability to identify optimal separating hyperplanes within the feature space. This improvement increased the model's precision and overall detection capabilities. This paper creates the use of an SVM-based method to classify the features derived from the fully trained MTAE at the final FC layer. To implement the SVM within the hybrid model, we utilize the Radial Basis Function (RBF) kernel. The penalty parameter, C , and the optimal kernel parameter, γ , are determined through validation data analysis on the hybrid model, which is trained using the training data.

The implementation of the MTAE-SVM model, as depicted in Figure 2, involves several key steps. First, during the training phase, the training set samples are input into the MTAE model. Once fully trained, the MTAE classifier automatically extracts the relevant features for each input. We then replace the FC layers with an SVM-based classifier, which learns from the extracted feature vectors produced by the MTAE. During testing, the fully trained MTAE processes the samples from the test set to generate feature representations. We then feed these extracted features into the trained SVM to detect PDs.

4. Experimental Setup

In this section, the datasets used to verify the performance our framework, feature extraction, and training setup, as well as performance criteria, are systematically presented.

4.1. Database

Two widely used databases for LS classification are the International Conference on Biomedical Health Informatics (ICBHI) and KAUH databases. In this paper, we employed the KAUH database to assess the effectiveness of our proposed methods. The primary reason for selecting this database was its suitability for subject-independent classification experiments based on fivefold cross-validation, as it includes more than four subjects per disease category (normal, asthma, pneumonia, COPD). The details of the KAUH database are shown in Table 1. The objective of this study was to assess the effectiveness of our proposed approach in identifying PDs from LS signals. The evaluation was conducted using a publicly available dataset, as cited in [30]. The dataset includes participants categorized into various health conditions: 35 classified as normal, 32 as asthma, 5 as pneumonia, and 9 as COPD. The LS signal was recorded at a sampling frequency of 4 kHz.

Table 1. The demographic information of the subjects based on the KAUH database.

Pulmonary Disease	No. of Subjects	Age Range	Gender (Female, Male)	Segmented Waveform No. (Original LS)
Normal	35	18–81	11 female, 24 male	415 (105)
Asthma	32	12–72	17 female, 15 male	357 (96)
Pneumonia	5	36–70	2 female, 3 male	66 (15)
COPD	9	42–76	1 female, 8 male	102 (27)

A crucial preprocessing stage in our methodology involved segmenting every LS signal into individual frames. Segmentation was accomplished by employing a non-overlapping window approach. In this study, we opted for window durations of 4 s due to their substantial impact on the overall detection performance of our method. Through extensive testing, the suitable window length was determined to ensure it captured essential clinical characteristics in the LS signals while optimizing the efficiency of the algorithm. The quantities of frames with a duration of 4 s are given in Table 1.

4.2. MFCC Configuration

The MFCC was utilized as the primary feature for inputting data into our model. The coefficients were calculated using frame lengths of 20, 40, 128, 256, 512, and 1000 ms. Each frame underwent processing with a Hamming window and had a 50% overlap. This approach resulted in a 13-dimensional MFCC. To investigate temporal changes in the audio frames, we computed both the delta and double-delta coefficients of the MFCC. Consistent with [40], we utilized a segment duration of 4 s in our investigation. Following the recommendation in [31], we converted the 39-dimensional coefficients of the CNN-2L-based classifier into a grayscale format with dimensions of 28×28 pixels. In contrast, our proposed model utilized the original 39-dimensional coefficients directly as input.

4.3. Network Training

The evaluation models were constructed using the PyTorch v1.10.1 framework and trained on an NVIDIA RTX3090 GPU with 24 GB of RAM. The training parameters can be found in Table 2.

Table 2. Model parameters of the MTAE.

Parameters	MTAE
Batch Size	32
Epochs	350
Optimizer	ADAM
Loss	Cross-entropy
Output Activation	Softmax
Hidden Layer Activation	ReLU

4.4. Performance Criteria

When assessing the classifier's performance, we employed commonly used metrics outlined in [41], described as precision (Pr), recall (Re), F1-score ($F1$), and accuracy (Acc), which can be defined as,

$$Pr = \frac{TP}{TP + FP} \quad (6)$$

$$Re = \frac{TP}{TP + FN} \quad (7)$$

$$F1 = 2 \times \frac{Pr \times Re}{Pr + Re} \quad (8)$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

where TP , TN , FP , and FN represent true positive, true negative, false positive, and false negative values, respectively.

5. Results and Discussion

This section showcases the results achieved by our proposed framework, alongside a comparative analysis of its performance with similar studies.

5.1. Results Based on MTAE Classification

This section unveils the outcomes of our proposed MTAE module. The subsequent subsections detail the findings obtained from adjusting the parameter configurations.

- The duration of the frame in the MFCC significantly influences the classification accuracy in deep learning models. Hence, identifying the optimal frame length for the MFCC is crucial to enhance the effectiveness of our proposed MTAE. This study employed an MTAE comprising three encoder and three decoder blocks for regression, with an additional three layers for classification. The architectural and configuration details of this MTAE are outlined in Table 3. A Learning Rate (LR) of 0.00001 was selected. The experimental results, illustrating the impact of various MFCC frame lengths on the performance of our MTAE, are presented in Table 4. As depicted in Table 4, the MFCC with a frame duration of 20 ms in our MTAE exhibited the highest efficacy in detecting PDs using LS signals. This selected frame length proved to be optimal as it strikes a balance between capturing spectral details and preserving temporal resolution, closely resembling the typical duration of a vowel sound in speech. This duration holds pivotal importance in LS studies, enabling precise collection of detailed acoustic characteristics necessary for distinguishing between normal and diseased sounds. Conversely, longer frame lengths tend to oversimplify these distinct characteristics, leading to a reduction in classification accuracy. This finding aligns with prior research presented in [28], underscoring its significance in LS studies for PD diagnosis. Consequently, we consistently adopted a frame length of 20 ms in the MFCC as the input parameter for our MTAE in all subsequent tests.
- Choosing the LR significantly influences the effectiveness of a deep learning model. An excessively high LR can cause the model to overshoot optimal weights, whereas an excessively low LR may lead to slow convergence or becoming stuck in suboptimal solutions. This study experimented with various LR values, and the results highlighted in Table 5 underscore the substantial impact of this hyperparameter. Remarkably, an LR of 0.0001 (LR = 0.0001) achieved the highest accuracy of 81.70%. This LR effectively adjusted the weights while preventing overshooting. Additionally, employing an LR of 0.001 (LR = 0.001) resulted in a commendable accuracy of 75.96%, demonstrating its suitability for the task at hand. However, setting the LR too high (0.01) or too low (0.000001) led to diminished model performance, underscoring the importance of careful LR adjustment for optimal training and convergence. Upon analyzing the results, we opted to utilize an LR of 0.0001 as the default setting for our MTAE in all subsequent studies, given its proven effectiveness in achieving optimal performance.
- The efficacy of our MTAE model is notably influenced by its architecture, particularly the configuration of encoder and decoder blocks during the regression phase, along with the number of FC layers in the classification component. The careful selection of these architectural elements plays a pivotal role in maximizing its effectiveness. We systematically conducted experiments by varying both the combinations of encoder-decoder blocks and the number of FC layers to assess their impact on accuracy. The outcomes of these adjustments, outlined in Tables 6 and 7, provide valuable insights into the optimal structural arrangement for our model. The results presented in Table 6 regarding encoder-decoder combinations reveal a significant relationship between model accuracy and architectural complexity. The configuration of 2-2 exhibits the highest accuracy rate of 89.47%, indicating that a well-balanced level of complexity is most effective for this model. This implies that having two layers in both the encoder and decoder achieves an optimal balance between capturing essential characteristics and preventing overfitting. In contrast, as the model's complexity increases with 3-3, 4-4, and 5-5 arrangements, there is a noticeable decline in accuracy. This decline is likely attributed to overfitting, where the model becomes overly attuned to the training data, thus compromising its ability to generalize. Remarkably, even the simplest 1-1 configuration still achieves a reasonably high accuracy of 88.51%. This suggests that

minimal configurations can be beneficial, although they are not as favorable as the 2-2 configuration. Based on the insights provided in Table 7, the highest accuracy is attained when utilizing three FC layers, achieving an accuracy of 89.47%. The second-highest accuracy, at 87.55%, is achieved with a configuration of two FC layers. This trend indicates that increasing the number of FC layers up to a certain threshold can enhance the model's data classification capabilities. However, the introduction of additional layers beyond this optimal threshold does not necessarily lead to performance improvement, as evidenced by the slight declines in accuracy observed for configurations with four and five FC layers (86.70% and 86.38%, respectively). This observation suggests the potential for overfitting or a limitation in learning capacity with additional layers. The model's performance is most compromised when employing only one FC layer, resulting in an accuracy of 85.84%. This underscores the necessity of multiple layers to achieve effective performance.

- To assess the effectiveness of multi-task learning, we adapted our model by removing the decoder component and relied solely on entropy loss for classification tasks. This modification led to a decrease in accuracy, with performance dropping to 88.40%. This outcome supports the hypothesis that a multi-task learning approach, utilizing a shared latent vector, is more efficient in detecting PDs from LS signals compared with a single-task learning technique that employs a non-shared latent vector. Such a technique aligns with similar methodologies outlined in the literature [27], suggesting that shared representations in multi-task learning may yield stronger and more adaptable models, particularly in complex tasks such as medical diagnosis.
- Following thorough parameter optimization, we settled on an LR of 0.0001 and finalized the design of our MTAE system. This architecture, featuring a 2-2 encoder-decoder arrangement and three FC layers, corresponds to the design depicted in Figure 2.

Furthermore, we evaluated the effectiveness of characteristics extracted from the trained MTAE in discerning various types of PDs in LSs. To accomplish this, we employed t-distributed Stochastic Neighbor Embedding (t-SNE) [42] for dimensionality reduction and comparison of raw sound signals, features extracted from the trained CNN-2L model, and features extracted from the learned MTAE model. We conducted an analysis of the distributions of four classes using t-SNE, selecting one out of five folds for each analysis. Figure 3 illustrates the t-SNE-based feature distributions obtained by comparing raw sound signals, features from the trained CNN-2L model, and features from the trained MTAE model. Specifically, Figure 3a showcases a noticeable convergence of class distributions when using raw sound signals, leading to limited distinction among diseases. In contrast, Figure 3b illustrates improved differentiation across classes using the CNN-2L model, albeit with some remaining overlap. Figure 3c demonstrates that features obtained via MTAE exhibit clearer boundaries and shorter gaps between different classes, indicating enhanced potential for effectively detecting PDs from LS signals.

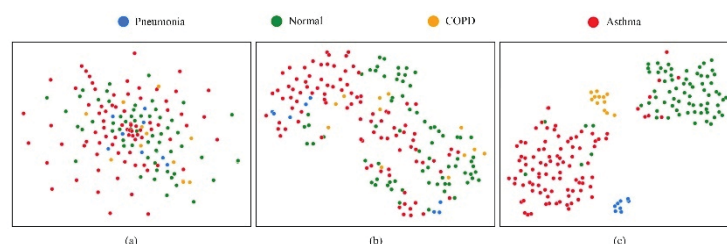


Figure 3. t-SNE visualization of feature distributions: (a) raw LS signal features, (b) features extracted from the baseline CNN model, and (c) features learned by the MTAE.

Lastly, to explore the relationship between each disease category, we present the confusion matrices. Figure 4 showcases the confusion matrices generated by our proposed framework. The confusion between COPD and asthma primarily stems from the similarities in their clinical presentations. However, our approach notably improves the categorization of pneumonia compared with the CNN-2L model, as evidenced by the increased true positive rate for pneumonia in matrix (b). The suggested MTAE exhibits a high overall accuracy in predicting the ‘normal’ class, while also mitigating the occurrence of false positives compared with the baseline. This suggests that the MTAE more effectively captures the unique characteristics of each disease, although it still encounters challenges in differentiating between diseases with similar symptoms.

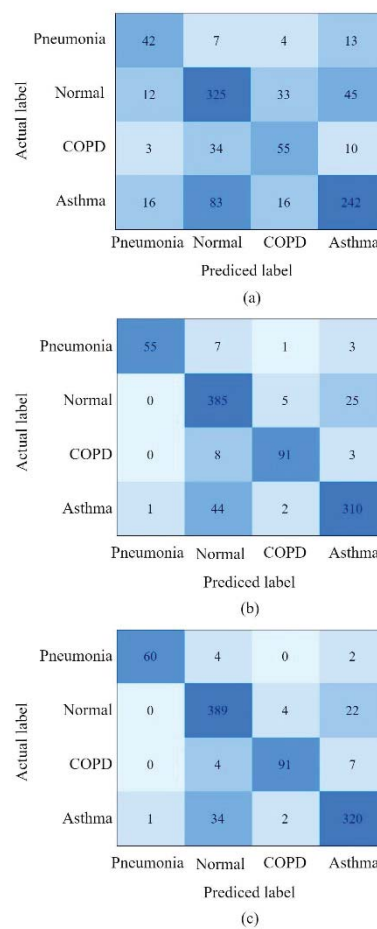


Figure 4. Confusion matrices for (a) CNN-based baseline, (b) MTAE, and (c) MTAE-SVM. The color intensity represents the proportion of correctly or incorrectly classified instances, with darker shades indicating higher values.

Table 3. Configuration of the MTAE designed for investigating the impact of varying frame lengths in MFCC analysis.

Module	Layer	Kemel	Layer Configuration	Activation
Auto. (Encoder1)	Conv2d 1	3×3	in_channels = 1, out_channels = 16	ReLU
Auto. (Encoder2)	Conv2d 2	3×3	in_channels = 16, out_channels = 8	ReLU
Auto. (Encoder3)	Conv2d 3	3×3	in_channels = 8, out_channels = 2	ReLU
Auto. (Decoder1)	ConvTranspose2d 1	3×3	in_channels = 2, out_channels = 8	ReLU
Auto. (Decoder2)	ConvTranspose2d 2	3×3	in_channels = 8, out_channels = 16	ReLU
Auto. (Decoder3)	ConvTranspose2d 3	3×3	in_channels = 16, out_channels = 1	ReLU and Sigmoid
Class. (Block1)	FC (Linear 1)	-	in_features = [flatten latent size], out_features = 128	ReLU
Class. (Block2)	FC (Linear 2)	-	in_features = 128, out_features = 128	ReLU
Class. (Block3)	FC (Linear 3)	-	in_features = 128, out_features = 128	ReLU
	Classifier (Linear 4)	-	in_features = 128, out_features = 4	Softmax

Auto.: autoencoder, Class.: classification.

Table 4. Effect of MFCC frame length on classification performance in the MTAE model.

Frame Length	Acc (%)	Pr (%)	Re (%)	F1 (%)
20 ms	71.28 $\pm$ 3.30	74.40 $\pm$ 9.21	52.60 $\pm$ 6.11	54.60 $\pm$ 5.13
40 ms	61.17 $\pm$ 5.10	47.60 $\pm$ 16.58	41.80 $\pm$ 5.50	40.20 $\pm$ 6.76
128 ms	56.60 $\pm$ 1.71	36.80 $\pm$ 11.34	35.40 $\pm$ 2.30	32.60 $\pm$ 2.70
256 ms	55.75 $\pm$ 4.62	33.00 $\pm$ 12.02	33.80 $\pm$ 2.28	31.20 $\pm$ 2.95
512 ms	50.74 $\pm$ 3.21	25.40 $\pm$ 2.61	30.80 $\pm$ 1.48	27.40 $\pm$ 1.52
1 s	51.38 $\pm$ 4.71	25.40 $\pm$ 2.51	31.40 $\pm$ 2.51	27.80 $\pm$ 2.49

Table 5. LR on MTAE-based classification.

LR	Acc (%)	Pr (%)	Re (%)	F1 (%)
0.01	44.15 $\pm$ 3.63	11.00 $\pm$ 0.71	25.00 $\pm$ 0.00	15.40 $\pm$ 0.89
0.001	75.96 $\pm$ 10.33	68.60 $\pm$ 20.50	67.20 $\pm$ 13.94	67.00 $\pm$ 17.59
0.0001	81.70 $\pm$ 3.44	80.20 $\pm$ 5.93	77.20 $\pm$ 3.56	78.00 $\pm$ 4.30
0.00001	71.28 $\pm$ 3.30	74.40 $\pm$ 9.21	55.60 $\pm$ 5.50	59.20 $\pm$ 7.33
0.000001	46.17 $\pm$ 5.30	18.40 $\pm$ 7.40	27.60 $\pm$ 3.21	21.40 $\pm$ 6.11

Table 6. Performance of different encoder-decoder configurations.

Encoder-Decoder	Acc (%)	Pr (%)	Re (%)	F1 (%)
1-1	88.51 $\pm$ 7.10	91.60 $\pm$ 5.13	89.00 $\pm$ 5.52	89.80 $\pm$ 5.72
2-2	89.47 $\pm$ 4.21	92.05 $\pm$ 3.15	88.72 $\pm$ 6.63	89.96 $\pm$ 5.07
3-3	81.70 $\pm$ 3.44	80.20 $\pm$ 5.93	77.20 $\pm$ 3.56	78.00 $\pm$ 4.30
4-4	72.87 $\pm$ 10.21	69.60 $\pm$ 14.84	63.80 $\pm$ 12.03	64.60 $\pm$ 12.28
5-5	71.17 $\pm$ 9.28	66.60 $\pm$ 9.66	65.60 $\pm$ 13.58	65.40 $\pm$ 11.63

Table 7. FC layer performance.

FC Layers	Acc (%)	Pr (%)	Re (%)	F1 (%)
1	85.84 $\pm$ 5.79	91.20 $\pm$ 2.77	84.60 $\pm$ 7.50	86.80 $\pm$ 5.36
2	87.55 $\pm$ 1.50	89.40 $\pm$ 2.88	85.80 $\pm$ 7.43	86.80 $\pm$ 4.44
3	89.47 $\pm$ 4.21	92.05 $\pm$ 3.15	88.72 $\pm$ 6.63	89.96 $\pm$ 5.07
4	86.70 $\pm$ 4.86	86.80 $\pm$ 5.12	84.20 $\pm$ 7.89	83.80 $\pm$ 9.18
5	86.38 $\pm$ 2.15	85.80 $\pm$ 3.63	82.20 $\pm$ 1.79	83.80 $\pm$ 1.48

5.2. Results Based on MTAE-SVM Classification

This subsection presents the results of the MTAE-SVM model. Since the γ value significantly affects SVM detection performance, identifying the optimal γ is essential.

Testing γ values from 0.1 to 30 in 0.1 increments, 0.1 emerged as optimal, offering the highest accuracy on the validation set of the trained hybrid model. The fully trained MTAE model with $\gamma = 0.1$ was then used on the test set because the distance between the MTAE-extracted features and the SVM hyperplane gave the best rate of correct classification. Figure 5 presents a comparison between the performance of the MTAE-SVM model and the MTAE model.

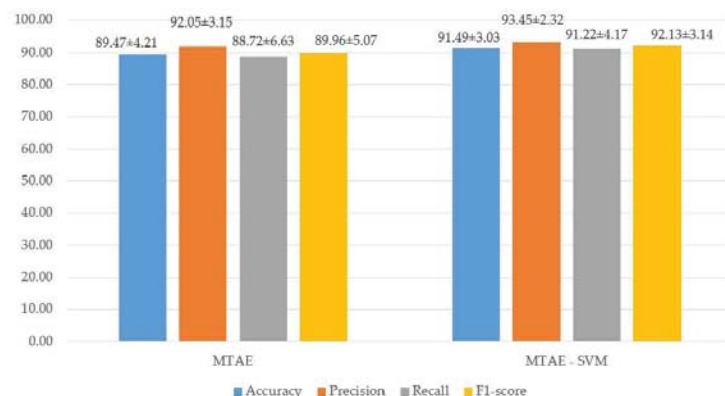


Figure 5. Comparison performance of the proposed MTAE classifier and MTAE-SVM classifier.

As shown in Figure 5, the hybrid model achieved enhanced accuracy, which can be attributed to its superior decision-making process. The increased distance between the automatically extracted features from the MTAE and the hyperplane of the trained SVM resulted in higher specificity compared with the MTAE classifier alone, as summarized in [28]. This directly contributed to improved detection accuracy, as evidenced by the greater detail in the MTAE-SVM confusion matrix compared with the MTAE and CNN methods, as illustrated in Figure 4a–c.

5.3. Comparison with Existing Approaches

In this subsection, we present a comprehensive comparison between our proposed feature extraction and classification technique and a range of established solutions. The evaluation is conducted by comparing the classification accuracy across four disease classes: normal (Acc_N), pneumonia (Acc_P), asthma (Acc_A), and COPD (Acc_C), as well as the Overall Accuracy (OA) of the system. The comparison of various feature extraction methods and classifiers for both the four-class and three-class scenarios is provided in Table 8 for the four-class case and in Table 9 for the three-class case.

Based on the information provided in Table 8, the proposed MTAE and MTAE-SVM achieve an OA of 89.47% and 91.49%, respectively, significantly outperforming the CNN 2L and EWT-FBP with LGBM, which attain OAs of 70.64% and 77.13%, respectively. These results highlight the superior capability of the MTAE and MTAE-SVM models in effectively capturing essential features for disease classification, leading to improved performance across multiple classes. The approach we propose demonstrates superior performance compared with alternative methods by utilizing the strengths of the MTAE's effective feature extraction alongside the strong classification abilities of the SVM. The MTAE model boosts feature learning through the simultaneous optimization of signal reconstruction and classification tasks, leading to the development of more distinct features. The incorporation of the SVM enhances classification performance by adeptly managing the representative features from the MTAE and optimizing the margin between classes, resulting in improved accuracy.

Table 8. Comparison with some systems for four classes.

Feature Extraction Method and Classifier	Acc_N (%)	Acc_P (%)	Acc_A (%)	Acc_C (%)	OA (%)
MFCC and CNN-2L, as in [31]	78.50	62.50	67.40	52.44	70.64
EWT-FBP and LGBM, as in [13]	-	-	-	-	77.13
MFCC and MTAE (ours)	92.77	83.33	86.83	89.22	89.47
MFCC and MTAE-SVM (ours)	93.73	90.91	89.64	89.22	91.49

Table 9. Comparison of feature extraction methods and classifiers for three classes.

Feature Extraction Method and Classifier	Acc_N (%)	Acc_P (%)	Acc_A (%)	OA (%)
MFCC and CNN-2L, as in [31]	85.30	40.01	64.79	72.31
MFCC and DNN-HMM, as in [24]	86.30	84.62	75.93	82.73
EWT-FBP and LGBM, as in [13]	77.58	100.00	76.72	84.76
MFCC and SVM, as in [24]	88.61	81.84	77.16	82.64
MFCC and KNN, as in [24]	76.67	72.22	66.67	72.05
MFCC and RF, as in [24]	80.28	79.27	59.26	73.96
Mel spectrogram and DNN, as in [24]	74.44	78.85	78.40	77.34
Mel spectrogram and LSTM, as in [24]	92.50	83.33	80.25	85.33
Mel spectrogram and GRU, as in [24]	92.22	81.41	78.70	83.94
Mel spectrogram & BiLSTM, as in [24]	94.44	84.62	79.63	86.28
Mel spectrogram & BiGRU, as in [24]	91.94	86.75	83.33	87.41
MFCC and MTAE (ours)	93.01	89.39	87.11	90.22
MFCC and MTAE-SVM (ours)	93.73	96.97	91.60	93.08

In the three-class classification task, as presented in Table 9, our proposed MTAE and MTAE-SVM systems achieved OAs of 90.22% and 93.08%, respectively, surpassing several well-known methods. The strong performance of the MTAE can be attributed to its multi-task learning approach, which effectively captures key MFCC features. Additionally, the hybrid methodology employed in the MTAE-SVM enhances performance, particularly in scenarios with limited training data, by fully leveraging the strengths of the MTAE. The impressive results of our models originate from their capacity to effectively learn complex patterns in LS signals and classify them with precision, a process further improved by the working together integration of the MTAE and SVM. This leads to enhanced precision in identifying PDs when compared with current methods.

6. Conclusions

In this paper, we introduced a novel MTAE model tailored specifically for the detection of PDs through the analysis of LS signals. The model employed a unique training approach that simultaneously enhances both classification accuracy and reconstruction quality. This model proposal has the potential to improve the efficiency and flexibility of audio-based medical diagnostics. The experimental findings demonstrate that our MTAE model outperforms CNN-2L and other existing systems in OA, achieving 89.47% accuracy in four-class disease classification and 90.22% accuracy in three-class classification. Furthermore, the MTAE-SVM could enhance the accuracy further. The accuracy could be improved to 91.49 % and 93.08 %, respectively. The remarkable outcome underscores the MTAE model's ability to capture the intricate and distinctive characteristics necessary for precise categorization of PDs. By fully using the strengths of the MTAE model, the hybrid approach of the MTAE-SVM also improves performance, especially when there are not many training data.

However, our study has certain limitations that need to be acknowledged:

- The model was evaluated using a specific dataset from KAUH, which may limit the generalizability of the results to other populations or recording conditions.
- The performance of the model relies on the availability of large numbers of labeled data for training, which can be challenging to obtain in medical settings.

- The model has not yet been implemented or validated in clinical settings, which is necessary to demonstrate its practical effectiveness in diagnosing PDs.

In future research direction, we plan to collaborate with healthcare institutions to test and refine the model in real-world scenarios, which will help demonstrate its potential in effectively diagnosing PDs. Additionally, we will explore the effectiveness of integrating phase feature extraction [43,44] and multi-scale CNNs [28] to improve our proposed MTAE methods. Furthermore, we will utilize score aggregation [45] to complement our suggested MTAE model with other classifiers.

Author Contributions: Conceptualization, K.O., K.P., W.P. (Watcharakorn Pinthurat), M.U. and P.U.; methodology, K.O.; Simulation, K.O.; validation, K.O., K.P., W.P. (Wongsathon Pathonsuwan), T.J., A.R., P.A., W.P. (Watcharakorn Pinthurat), M.U. and P.U.; visualization, K.O. and K.P.; formal analysis, K.O., K.P. and W.P. (Watcharakorn Pinthurat); writing—original draft, K.O.; writing—review and editing, K.O., K.P., W.P. (Wongsathon Pathonsuwan), T.J., A.R., P.A., W.P. (Watcharakorn Pinthurat), M.U. and P.U.; supervision, K.P., M.U. and P.U. All authors have read and agreed to the published version of the manuscript.

Funding: This research project is supported by Science Research and Innovation Fund. Agreement No. FF67/P1-044.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study were collected from King Abdullah University Hospital and are openly available in [Mendeley Data] at [<https://doi.org/10.17632/jwyy9np4gv.3> accessed on 9 September 2024], reference number [30].

Acknowledgments: This research project is supported by Suranaree University of Technology (SUT), Thailand.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Kazem, Z.; Zahra, M. Fundamentals of respiratory sounds and analysis. *Morgan Claypool Publ. Ser. Synth. Lect. Biomed. Eng.* **2006**, *8*, 19–21.
2. Gibson, G.J.; Loddenkemper, R.; Lundbäck, B.; Sibille, Y. Respiratory health and disease in Europe: The new European Lung White Book. *Eur. Respir. J.* **2013**, *42*, 559–563. [CrossRef] [PubMed]
3. Shi, L.; Zhang, Y.; Zhang, J. Lung sound recognition method based on wavelet feature enhancement and time-frequency synchronous modeling. *IEEE J. Biomed. Health Inform.* **2022**, *27*, 308–318. [CrossRef]
4. Shuvo, S.B.; Ali, S.N.; Swapnil, S.I.; Hasan, T.; Bhuiyan, M.I.H. A lightweight CNN model for detecting respiratory diseases from lung auscultation sounds using EMD-CWT-based hybrid scalogram. *IEEE J. Biomed. Health Inform.* **2020**, *25*, 2595–2603. [CrossRef]
5. Abbasi, S.; Derakhshanfar, R.; Abbasi, A.; Sarbaz, Y. Classification of normal and abnormal lung sounds using neural network and support vector machines. In Proceedings of the 21st Iranian Conference on Electrical Engineering (ICEE), Mashhad, Iran, 14–16 May 2013; pp. 1–4.
6. Uysal, S.; Uysal, H.; Bolat, B.; Yildirim, T. Classification of normal and abnormal lung sounds using wavelet coefficients. In Proceedings of the Signal Processing and Communications Applications Conference (SIU), Trabzon, Turkey, 23–25 April 2014; pp. 2138–2141.
7. Liu, Y.; Zhang, C.-M.; Zhao, Y.H.; Dong, L. The feature extraction and classification of lung sounds based on wavelet packet multiscale analysis. *Chin. J. Comput.-Chin. Ed.* **2006**, *29*, 769.
8. Sengupta, N.; Sahidullah, M.; Saha, G. Lung sound classification using cepstral-based statistical features. *Comput. Biol. Med.* **2016**, *75*, 118–129. [CrossRef]
9. Jakovljević, N.; Lončar-Turukalo, T. Hidden Markov model-based respiratory sound classification. In Proceedings of the Precision Medicine Powered by pHealth and Connected Health: ICBHI 2017, Thessaloniki, Greece, 18–21 November 2017; pp. 39–43.
10. Bahoura, M.; Pelletier, C. New parameters for respiratory sound classification. In Proceedings of the CCECE 2003-Canadian Conference on Electrical and Computer Engineering. Toward a Caring and Humane Technology, Montreal, QC, Canada, 4–7 May 2003; Volume 3, pp. 1457–1460.
11. Jung, S.-Y.; Liao, C.-H.; Wu, Y.-S.; Yuan, S.-M.; Sun, C.-T. Efficiently classifying lung sounds through depthwise separable CNN models with fused STFT and MFCC features. *Diagnostics* **2021**, *11*, 732. [CrossRef]

12. Azmy, M.M. Classification of lung sounds based on linear prediction cepstral coefficients and support vector machine. In Proceedings of the IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), Amman, Jordan, 3–5 November 2015; pp. 1–5.
13. Tripathy, R.K.; Das, D.; Ghose, A.; Ghosh, A. Automated detection of pulmonary diseases from lung sound signals using fixed-boundary-based empirical wavelet transform. *IEEE Sens. Lett.* **2022**, *6*, 1–4. [\[CrossRef\]](#)
14. Prajapati, S.K.; Choudhary, T.S.; Mishra, S. Early Detection of Lung Disease Using Multi-Class Classifiers. In Proceedings of the 2023 IEEE 4th Annual Flagship India Council International Subsections Conference (INDISCON), Mysore, India, 5–7 August 2023; pp. 1–6.
15. Reynolds, D.A.; Rose, R.C. Robust text-independent speaker identification using Gaussian mixture speaker models. *IEEE Trans. Audio Speech Lang. Process.* **1995**, *3*, 72–83. [\[CrossRef\]](#)
16. Haider, A.; Ashraf, M.D.; Azhar, M.U.; Maruf, S.O.; Naqvi, M.; Khawaja, S.G.; Akram, M.U. Separation and classification of crackles and bronchial breath sounds from normal breath sounds using Gaussian mixture model. In Proceedings of the Neural Information Processing: 21st International Conference, Kuching, Malaysia, 3–6 November 2014; pp. 495–502.
17. Fraiwan, L.; Hassanin, O.; Fraiwan, M.; Khassawneh, B.; Ibnian, A.M.; Alkhodari, M. Automatic identification of respiratory diseases from stethoscopic lung sound signals using ensemble classifiers. *Biocybern. Biomed. Eng.* **2021**, *41*, 1–14. [\[CrossRef\]](#)
18. Chambres, G.; Hanna, P.; Desainte-Catherine, M. Automatic detection of patients with respiratory diseases using lung sound analysis. In Proceedings of the International Conference on Content-Based Multimedia Indexing (CBMI), La Rochelle, France, 4–6 September 2018; pp. 1–6.
19. Jaber, M.M.; Abd, S.K.; Shakeel, P.M.; Burhanuddin, M.A.; Mohammed, M.A.; Yusof, S. A telemedicine tool framework for lung sounds classification using ensemble classifier algorithms. *Measurement* **2020**, *162*, 107883. [\[CrossRef\]](#)
20. García-Ordás, M.T.; Benítez-Andrades, J.A.; García-Rodríguez, I.; Benavides, C.; Alaiz-Moreton, H. Detecting respiratory pathologies using convolutional neural networks and variational autoencoders for unbalanced data. *Sensors* **2020**, *20*, 1214. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Wanasinghe, T.; Bandara, S.; Madusanka, S.; Meedeniya, D.; Bandara, M.; de la Torre Díez, I. Lung Sound Classification With Multi-Feature Integration Utilizing Lightweight CNN Model. *IEEE Access* **2024**, *12*, 21262–21276. [\[CrossRef\]](#)
22. Bardou, D.; Zhang, K.; Ahmad, S.M. Lung sounds classification using convolutional neural networks. *Artif. Intell. Med.* **2018**, *88*, 58–69. [\[CrossRef\]](#)
23. Li, L.; Xu, W.; Hong, Q.; Tong, F.; Wu, J. Classification between normal and adventitious lung sounds using deep neural network. In Proceedings of the International Symposium on Chinese Spoken Language Processing (ISCSLP), Tianjin, China, 17–20 October 2016; pp. 1–5.
24. Shi, L.; Du, K.; Zhang, C.; Ma, H.; Yan, W. Lung sound recognition algorithm based on VGGish-BIGRU. *IEEE Access* **2019**, *7*, 139438–139449. [\[CrossRef\]](#)
25. Dithaporn, A.; Banluesombatkul, N.; Ketrat, S.; Chuangsuwanich, E.; Wilaiprasitporn, T. Universal joint feature extraction for P300 EEG classification using multi-task autoencoder. *IEEE Access* **2019**, *7*, 68415–68428. [\[CrossRef\]](#)
26. Autthasan, P.; Chaisaen, R.; Sudhawiyangkul, T.; Rangpong, P.; Kiatthaveephong, S.; Dilokthanakul, N.; Bhakdisongkhram, G.; Phan, H.; Guan, C.; Wilaiprasitporn, T. MIN2Net: End-to-end multi-task learning for subject-independent motor imagery EEG classification. *IEEE Trans. Biomed. Eng.* **2021**, *69*, 2105–2118. [\[CrossRef\]](#)
27. Bashir-Gonbadi, F.; Khotanlou, H. Brain tumor classification using deep convolutional autoencoder-based neural network: Multi-task approach. *Multimed. Tools Appl.* **2021**, *80*, 19909–19929. [\[CrossRef\]](#)
28. Pathonsuwan, W.; Phapatanaburi, K.; Buayai, P.; Jumphoo, T.; Anchuen, P.; Uthansakul, M.; Uthansakul, P. RS-MSConvNet: A novel end-to-end pathological voice detection model. *IEEE Access* **2022**, *10*, 120450–120461. [\[CrossRef\]](#)
29. Niu, X.-X.; Suen, C.Y. A novel hybrid CNN-SVM classifier for recognizing handwritten digits. *Pattern Recognit.* **2012**, *45*, 1318–1325. [\[CrossRef\]](#)
30. Fraiwan, M.; Fraiwan, L.; Khassawneh, B.; Ibnian, A. A dataset of lung sounds recorded from the chest wall using an electronic stethoscope. *Data Br.* **2021**, *35*, 106913. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Aykanat, M.; Kılıç, Ö.; Kurt, B.; Saryal, S. Classification of lung sounds using convolutional neural networks. *Eurasip J. Image Video Process.* **2017**, *2017*, 1–9. [\[CrossRef\]](#)
32. Nukitram, J.; Saengmolee, W.; Chaisaen, R.; Autthasan, P.; Sengnon, N.; Wungsintaweekul, J.; Cheaha, D.; Kumarnsit, E.; Sudhawiyangkul, T.; Wilaiprasitporn, T. ANet: Autoencoder-Based Local Field Potential Feature Extractor for Evaluating an Antidepressant Effect in Mice After Administering Kratom Leaf Extracts. *IEEE Trans. Biomed. Circuits Syst.* **2023**, *17*, 67–76. [\[CrossRef\]](#)
33. Das, B.B.; Ram, S.K.; Babu, K.S.; Mohapatra, R.K.; Mohanty, S.P. Person Identification Using Autoencoder-CNN Approach with Multitask-Based EEG Biometric. *Multimed. Tools Appl.* **2024**, *83*, 83205–83225.
34. Li, Y.; Gong, S.; Liddell, H. Support vector regression and classification based multi-view face detection and recognition. In Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, Grenoble, France, 28–30 March 2000; pp. 300–305.
35. Toledo-Pérez, D.C.; Rodríguez-Reséndiz, J.; Gómez-Loenzo, R.A.; Jauregui-Correa, J.C. Support vector machine-based EMG signal classification techniques: A review. *Appl. Sci.* **2019**, *9*, 4402. [\[CrossRef\]](#)

36. Ismail, A.; Abdlerazek, S.; El-Henawy, I.M. Development of smart healthcare system based on speech recognition using support vector machine and dynamic time warping. *Sustainability* **2020**, *12*, 2403. [\[CrossRef\]](#)
37. Khairandish, M.O.; Shamma, M.; Jain, V.; Chatterjee, J.M.; Jhanjhi, N.Z. A hybrid CNN-SVM threshold segmentation approach for tumor detection and classification of MRI brain images. *Irbm* **2022**, *43*, 290–299. [\[CrossRef\]](#)
38. Liu, G.; Wang, L.; Liu, D.; Fei, L.; Yang, J. Hyperspectral image classification based on non-parallel support vector machine. *Remote Sens.* **2022**, *14*, 2447. [\[CrossRef\]](#)
39. Lin, C.-C.; Chen, S.-H.; Truong, T.-K.; Chang, Y. Audio classification and categorization based on wavelets and support vector machine. *IEEE Trans. Speech Audio Process.* **2005**, *13*, 644–651.
40. Sengupta, N.; Sahidullah, M.; Saha, G. Optimization of cepstral features for robust lung sound classification. In Proceedings of the IEEE India Conference (INDICON), New Delhi, India, 17–20 December 2015; pp. 1–6.
41. Thuwajit, P.; Rangpong, P.; Sawangjai, P.; Autthasan, P.; Chaisaen, R.; Banluesombatkul, N.; Boonchit, P.; Tatsaringkansakul, N.; Sudhawiyangkul, T.; Wilaiprasitporn, T. EEGWaveNet: Multiscale CNN-based spatiotemporal feature extraction for EEG seizure detection. *IEEE Trans. Ind. Inform.* **2021**, *18*, 5547–5557. [\[CrossRef\]](#)
42. Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
43. Phapatanaburi, K.; Pathonsuwan, W.; Wang, L.; Anchuen, P.; Jumphoo, T.; Buayai, P.; Uthansakul, M.; Uthansakul, P. Whispered speech detection using glottal flow-based features. *Symmetry* **2022**, *14*, 777. [\[CrossRef\]](#)
44. Phapatanaburi, K.; Wang, L.; Liu, M.; Nakagawa, S.; Jumphoo, M.T.; Uthansakul, M. Significance of relative phase features for shouted and normal speech classification. *EURASIP J. Audio Speech Music Process.* **2024**, *1*, 2. [\[CrossRef\]](#)
45. Phapatanaburi, K.; Buayai, P.; Kupimai, M.; Yodrot, T. Linear prediction residual-based constant-Q cepstral coefficients for replay attack detection. In Proceedings of the International Electrical Engineering Congress (iEECON), Chiang Mai, Thailand, 4–6 March 2020; pp. 1–4.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.