Classification of Five Lung Sound Types Using STVMD and GLCM-GLDM-based Feature Extractors

Annisa Humairani [1,2,*], Achmad Rizal [1,2], Salsabila Aurellia [1,2], and Rizkyandito A. Saputra [1]

[1] School of Electrical Engineering, Telkom University, Bandung, Indonesia
[2] Center of Excellence for Biomedical and Healthcare Technology, Research Institute for Digital Health, Social and Wellness, Telkom University, Bandung, Indonesia
Email: annisahumairani@telkomuniversity.ac.id (A.H.); achmadrizal@telkomuniversity.ac.id (A.R.); salsabilaaurelliasa@telkomuniversity.ac.id (S.A.); rizkyandito@student.telkomuniversity.ac.id (R.A.S.)
*Corresponding author

***Abstract*—Lung auscultation is fast and low-cost, but its interpretation is subjective and sensitive to recording variability, leading to unidentified cases and elevated fatality rates. Automated screening is therefore necessary. Accurate identification of lung sound types remains challenging due to subtle acoustic variations and overlapping characteristics. The methods developed from current research rely on features produced by the Time Frequency Representations (TFR), which impact the discriminability and robustness. This study proposes Short-Time Variational Mode Decomposition (STVMD), a combination of the Short-Time Fourier Transform (STFT) with Variational Mode Decomposition (VMD) in two Strategies (STGs). Strategy 1 (STG 1) applies VMD-2D to the spectrograms, targeting the enhancement of 2D texture structures. While Strategy 2 (STG 2) applies VMD-1D to the waveform before spectrogram generation to improve TFR concentration through the pre-separation of 1D components. Gray-Level Co-occurrence Matrix (GLCM) and Gray-Level Difference Matrix (GLDM) features are extracted to be classified using a Support Vector Machine (SVM). On a five-class lung sound dataset, the method achieved the highest accuracy at 97.98% for STG 1 with substantially lower feature extraction time, 0.16 min. These results show that STG 1 offers the best accuracy-efficiency trade-off, enhancing discriminative spectrogram textures for GLCM-GLDM at minimal computational cost.**

***Keywords*—auscultation, lung sound classification, short-time Fourier transform, variational mode decomposition**

I. Introduction

A condition of Lung Disease (LD) is any pulmonary disorder that interferes with normal respiratory system functioning. Generally, there are three types of LD, which are airway, lung tissue, and circulation [1]. These types involve the severity of shortness of breath and lead to changes in lung sounds, which are associated with the extent of pulmonary injury [2]. Since the impairments directly compromise ventilation and gas exchange, lung diseases collectively rank among the leading causes of death worldwide. There have been around 262 million people affected by asthma, including 455,000 deaths [3, 4]. Meanwhile, Chronic Obstructive Pulmonary Disease (COPD), which is often linked to bronchial-type airflow limitation and wheeze, caused 3.5 million deaths in 2021 [5]. Furthermore, there are Lower Respiratory Infections (LRIs), notably pneumonia, pulmonary embolism, and other infections often associated with crackles and a pleural rub that is generated from pleural inflammation [6, 7]. LRIs continue to impose a substantial population-level burden with estimated global incident cases and deaths around 488.9 and 2.4 million, respectively [6]. In addition, upper-airway obstruction may also produce stridor [7]. These lung sounds provide clinicians with key cues for identifying abnormalities in the lungs and broader respiratory tract [8]. Therefore, early-stage screening is needed to reduce the number of cases and mortality.

In spite of major developments in imaging-based diagnostics, cost and access barriers can limit routine use in resource-limited settings. Concurrently, non-invasive lung auscultation remains rapid, but its interpretation is often done subjectively and is vulnerable the interference from heart sounds, noise, and inter-observer variability. These constraints have motivated the development of Digital Signal Processing (DSP) and Machine Learning (ML) methods to standardize auscultation and automate lung sound classification [9–11].

Recent surveys highlight that Time-Frequency Representation (TFR) significantly affects the initial pipeline of lung sound classification, such as Short-Time Fourier Transform (STFT), Scalogram, Mel-spectrogram, Mel-Frequency Cepstral Coefficients (MFCC), or Gammatone gram [12]. TFRs' advantage is that the unique patterns of lung sounds are provided through a visual representation of the frequency spectrum of a

Manuscript received March 9, 2026; revised May 11, 2026; accepted June 30, 2026; published November 14, 2026.

sound signal. These TFRs are typically fed directly into conventional classifiers or deep learning models after the pre-processing stage [10]. Although the recordings are clean from noise, lung sounds often reveal overlapping continuous and transient components, producing mixed time-frequency patterns that are not explicitly mode-separated by a single TFR [13]. Thus, a key gap is that most methods treat the TFR as a final representation, without decomposing it into mode-relevant components to enhance class separability and representation stability.

Although decomposition has been studied to reduce signal complexity and improve separability, the algorithms most frequently used are wavelet-based or Empirical Mode Decomposition (EMD) approaches. These methods are applied to the signal prior to feature extraction and classification rather than explicitly decomposing the TFR itself [14–16]. Their performances can be sensitive to the choice of the parameters, mode mixing, and inconsistent mode structure across samples, which escalates the complexity for feature extraction and classification processes [17]. Variational Mode Decomposition (VMD), in contrast, formulates decomposition as a constrained variational optimization that extracts a set of narrow-band components with specific center frequencies while preserving accurate signal reconstruction. VMD has shown promise for lung sound processing in extracting the adventitious components, but remains underexplored in multi-class lung sound classification [18].

In line with these directions, this study proposed a Short-Time Variational Mode Decomposition (STVMD)-based lung sound classification through Strategy 1 (STG 1) and Strategy 2 (STG 2). In STG 1, the process starts by computing the STFT to obtain TFRs. Then, applies a VMD-2D-based decomposition perspective from the TFR results to separate mixed structures into mode-relevant components. In STG 2, as compared to the process, it measures VMD-1D-based decomposition as the input of STFT to result the TFRs, indicating the reverse of the processing order. Both orderings are evaluated since VMD can operate either on the raw 1D signal or 2D TFR. These choices affect the type of structures emphasized for the downstream features that are calculated using Gray-Level Co-occurrence Matrix and Gray-Level Difference Matrix (GLCM-GLDM) texture features to capture both co-occurrence and intensity-difference patterns. The features are classified using a multi-class Support Vector Machine (SVM) among the five target lung sound categories. This study seeks to accurately classify lung sounds into five categories for reliable screening support. The main contributions are as follows:

- Introduce an STFT-to-VMD decomposition strategy (STG 1) for lung sound analysis, where VMD-2D is applied from the STFT-derived TFR perspective to separate mixed TFR structures into mode-relevant components. This algorithm enables more discriminative representations to improve the classification performance considering that the TFR textures are stably enhanced. For comparison, as compared strategy with VMD-to-STFT decomposition to reduce smearing by feeding STFT band-limited modes.
- Utilize a lightweight feature extraction scheme using GLCM-GLDM to provide compact representations and interpretations of the decomposed time-frequency textures for downstream classification.
- Demonstrate accurate classification using a multi-class SVM classifier to categorize recordings into five lung sound classes (normal, wheeze, crackle, pleural rub, and stridor), which is critical for reliable diagnosis and informed treatment planning.

The remainder of this paper is organized as follows: Section II elaborates on the literature review of existing methods for lung sound classification; Section III shows the detailed steps and proposed methods for STVMD-based lung sound classification; Section IV provides the results and discussion; and Section V concludes the overall study.

II. Literature Review

In this section, a literature review is presented and organized around the latest lung sound classification using various datasets, highlighting its advantages and challenges. The shortcomings of previous methods are explained to enhance the development of more effective classification models.

In 2020, Ma *et al.* [19] proposed an improved adventitious lung sound classification model using LungRN + NL on the International Conference on Biomedical and Health Informatics (ICBHI) 2017 dataset. LungRN + NL was a ResNet-based model enhanced with a non-local layer and mix-up augmentation. STFT was used as the TFR inputs, and the non-local block was designed to model dependencies across time and frequency regions. The method achieved a low performance score classifying normal, crackle, wheeze, and crackle plus wheeze, leaving a high potential in developing the lung sound classification [19].

In a related study using a smaller dataset of 76 patients in 2022, Kim *et al*. [20] implemented Mel-spectrogram as the TFR method to extract the audio features and classified using ResNet34 with Convolutional Block Attention Module (CBAM) model. Mel-spectrogram features were advantageous to uncover valuable information about the frequency components and their strength from the waveform inputs. In classifying wheeze sounds from other lung sounds, they achieved a higher result than [19], showing that TFR could perform better with the right methodology [20].

In the same year, 2022, Rizal *et al*. [21] demonstrated pulmonary crackles and normal lung sounds classification applying STFT with multiple resolutions. The model was tested on the dataset collected from several sources, including 21 and 22 pulmonary crackles and normal lung sounds, respectively. The results were excellent, enabling a higher potential of STFT-based features for lung sound classification. The limitation of this research is not only that they classify two classes, but also that the number of

features used was relatively higher than in other compared studies [21].

Following the research, Dubey *et al.* [22] classified lung sounds using AI, specifically SVM and Long Short-Term Memory (LSTM), in 2023. The workflow starts with pre-processing signals from the RALE® database, then extracts the features using Wavelet Bi-Spectrum (WBS) and Bi-Phase (WBP). The application of WBP and WBS was to improve the classification process by characterizing the non-linearities of adventitious sounds. The classification was performed on three classes, i.e., normal, crackle, and wheeze. Although WBS with LSTM gave satisfactory results, more improvements can still be executed to improve the reliability of automated lung sound classification [22].

Further investigating in 2024, Orkweha *et al.* [23] introduced a combination of Multi-Task Autoencoder models and Support Vector Machine (MTAE-SVM) with 5-fold Cross-Validation (CV) to optimize performance. MFCC, with several frame lengths, serves as the primary input to the classifiers. Their experiments were tested on the King Abdullah University Hospital (KAUH) lung sound database containing 81 participants distributed across four conditions, which are normal, asthma, pneumonia, and COPD. Their algorithm obtained stable results in classifying 3 and 4 classes. However, the study highlights the ongoing difficulty of separating acoustically similar respiratory conditions when using compact MFCC-driven pipelines on limited-subject datasets [23].

More recently, in early 2026, Engin *et al.* [24] examined the effect of applying different TFRs to classify common lung sounds using several CNN features. The lung sound dataset was recorded first from 94 individuals who visited the Department of Chest Diseases at Karadeniz Technical University. There were four classes classified, which are normal, rhonchi, fine crackle, and coarse crackle. The highest result was achieved by Gammatone gram TFR combined with DenseNet201. This study indicates that TFRs were useful to analyze lung sounds since it has the ability to capture subtle spectral details in lung sounds more effectively. However, their study depends on accurate auto-detection of respiratory-cycle boundaries, so segmentation errors can propagate and degrade classification performance [24].

III. Materials and Methods

The main focus of this study is STVMD deployment prior to feature extraction and classification for the lung sound signals. As depicted in Fig. 1, the lung sound recordings from the lung sound dataset were first pre-processed to improve signal quality and standardize input conditions. Subsequently, two alternative STVMD strategies were constructed. Strategy 1 (STG 1) generated an STFT-based TFR and then applied VMD-2D to decompose the TFR into mode-relevant components. Strategy 2 (STG 2) applied VMD-1D to the time-domain signal, followed by STFT to form a mode-specific TFR. STG 1 targets the enhancement of 2D texture, whereas STG 2 focuses on the pre-separation of 1D components to improve TFR concentration. From each strategy's decomposed representation, GLCM and GLDM features were extracted to capture spatial dependency and intensity-variation characteristics. Finally, a multi-class SVM used for classification, resulting in several metric evaluations, such as accuracy, precision, sensitivity, and F1-Score.

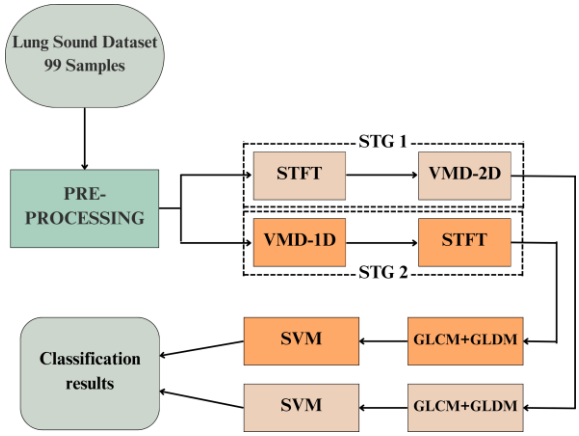


Fig. 1. Block diagram of the proposed STVMD-based lung sound classification framework.

A. Lung Sound Dataset

The dataset used in this study is available under the name 5-Class Lung Sounds Dataset from Several Sources published on Kaggle [25, 26]. Previously, this dataset was already used in earlier studies to classify the applicable classes [21, 27]. There are five classes provided, representing normal and abnormal lung sounds with details shown in Table I. The dataset consists of 99 recordings that were collected from patients aged 2 weeks to 79 years. Accordingly, the recordings capture a broad range of auscultation characteristics, consisting of both Continuous Adventitious Sounds (CAS) and Discontinuous Adventitious Sounds (DAS). The dataset consists of WAV recordings segmented to one full breathing cycle, inspiration-expiration, sampled at 8000 Hz with lengths of about 15,000–34,000 samples. The heart-sound contamination is none in the recordings because it has been removed in the previous study since explicit heart-sound removal remains outside the scope of this work [27, 28]. Moreover, the dataset provides balanced data of each class as seen in Table I [15].

It is important to note that this study utilizes a compilation dataset rather than a single standardized benchmark. The ICBHI-2017 dataset is the most widely used public benchmark in automated lung sound classification. Nonetheless, it encompasses only four classes, which are normal, crackle, wheeze, and crackle-plus-wheeze, and does not include pleural rub and/or stridor. The dataset used in this study extends the classification scope to five classes, enabling evaluation of a broader and more clinically diverse set of adventitious sounds. Moreover, the use of this dataset reflects a deliberate design choice to address a more comprehensive classification problem.

TABLE I. DETAILS OF CLASS DISTRIBUTION IN THE DATASET

Class	Total Number of Recordings (Data)	Percentage (%)
Normal	22	22.22
Wheeze	18	18.18
Crackle	21	21.21
Pleural Rub	18	18.18
Stridor	20	20.20
Total	**99**	**100**

Normal bronchial sounds are ordinary lung sounds characterized by a longer expiratory phase and a distinct pause between inspiration and expiration. Wheeze is a continuous "musical" adventitious sound, often high-pitched, that may be heard during inspiration, expiration, or both. On the contrary, crackles are non-musical, brief, explosive, discontinuous sounds, commonly prominent during inspiration. Pleural rub is a non-rhythmic adventitious sound generated when inflamed pleural surfaces rub during breathing, and it may be audible in both inspiratory and expiratory phases. The last sound from the dataset is stridor, which is a harsh, high-pitched sound associated with upper-airway narrowing, and often most evident on inspiration [15].

Each lung sound class in this dataset is clinically associated with specific pulmonary conditions. Normal class is characteristic of healthy respiratory function. Wheeze is commonly observed in obstructive airway diseases such as asthma, COPD, and bronchitis [5]. Crackles are frequently linked to conditions involving fluid accumulation or fibrosis, including pneumonia, heart failure, and idiopathic pulmonary fibrosis [6]. Pleural rub is a hallmark of pleuritis or pleurisy, and may also appear in pulmonary embolism and pneumonia when pleural inflammation is present. Stridor is most often associated with upper-airway obstruction, including croup, epiglottitis, and foreign body aspiration [7]. Recognizing these clinical associations is important for contextualizing the classification task and for eventually linking automated screening outputs to actionable clinical decisions.

B. Pre-processing

The lung sound signals were pre-processed using mean removal and amplitude normalization. As defined in Eqs. (1) and (2), subtracting the sample mean and scaling by the maximum absolute amplitude yields a zero-mean signal with values constrained to the range [−1, +1]. This step reduces variability introduced by differing recording conditions [29].

$$LS_0(n) = LS_i(n) - \frac{1}{N}\sum_{i=1}^{N} LS_i(i) \tag{1}$$

$$LS_0(n) = \frac{LS_i(n)}{\max\left|LS_i(i)\right|} \tag{2}$$

where $LS_0(n)$ is the pre-processed signal, $LS_i(n)$ is the lung sound input signal, and n is the signal length.

C. Signal Decomposition

In the signal decomposition stage, we employed STFT and VMD to capture complementary representations of lung sound signals. The resulting decomposed components served as the basis for subsequent feature extraction and classification.

1) Short-Time Fourier Transform (STFT)

The principle of STFT is the computation of the Fast Fourier Transform (FFT) of each data frame, obtained from a signal that is windowed into a fixed time window or short consecutive data frames. The process of obtaining data frames starts with multiplying a signal by a window function of length. The window length not only determines the frequency resolution, but also the time resolution of the signal in the spectrum. Then, there is the window type which helps to decide the time-frequency distribution [30]. FFT is performed using Eq. (3) below.

$$X_{STFT}(n,\omega) = \sum_{\infty}^{\infty} X_i(i)\, w(n-i)\, e^{j\omega n} \tag{3}$$

where $X_i(i)$ is the input signal and $w(i)$ is the window function that centred at time n.

In this study, two types of input signals were considered, viz. the pre-processed and decomposed signal. Kaiser window function was applied with window/overlap length pairs of 25/20 and 200/100, and FFT sizes of 256 and 512 points [31]. The analysis using STFT was used for generating TFR as 2D images from the 1D raw and decomposed lung sound signals from the dataset. All resulting STFT magnitude spectrograms were standardized to a fixed 128 × 128 TFR. Then, the variations of the frequency distribution over time were described by the visual information in a spectrogram [12, 21].

2) VMD-1D and VMD-2D

VMD is an adaptive signal decomposition technique that formulates the extraction of oscillatory components as a constrained variational optimization problem. In VMD-1D, the raw 1D signal is decomposed into several sub-signals, popularly known as modes, iteratively. Every decomposed mode is localized in the frequency domain around a centre frequency.

The Intrinsic Mode Functions (IMFs) and their centre frequencies are estimated through an optimization algorithm. Specifically, the algorithm minimizes the total bandwidth across modes while enforcing that their sum reconstructs the original signal, exploiting the Alternating Direction Method of Multipliers (ADMM). The optimization algorithm is mathematically written as Eq. (4) below [32].

$$x = \min_{x_m,\omega_m}\left\{\sum_m \left\|\partial_t\left[\left(\delta(t)+\frac{j}{\pi t}\right)\cdot x_m(t)\right]e^{-j\omega_m t}\right\|_2^2\right\} s.t. \sum_m x_m \tag{4}$$

where m is the mode number, $x_m(t)$ is the signal in the related mode with centre frequency ω_m, δ is Dirac delta,

∂_t denotes the partial derivative with respect to time, and t is time.

The mathematical expression resulted from some steps which starts by the unilateral frequency spectrum signal computation using the means of the Hilbert transform. Next, the frequency spectrum is shifted to the baseband by multiplying an exponential signal with the estimated centre frequency. Lastly, the bandwidth is measured through $H1$ Gaussian smoothness of the demodulated signal, defined as the squared $L2$-norm of its gradient.

As for VMD-2D decomposition, the estimation is formulated in Eqs. (5) and (6) below.

$$X^{n+1}(\omega) \leftarrow \frac{X(\omega) - \sum_{i<m} X_i^{n+1} - \sum_{i<m} X_i^n + \frac{\lambda^n}{2}}{1 + 2\alpha\left(\omega - \omega_m^n\right)^2} \tag{5}$$

$$\omega_m^{n+1} \leftarrow \frac{\int_0^\infty \omega \left| X_k^{n+1}(\omega) \right|^2 d\omega}{\int_0^\infty \left| X_k^{n+1}(\omega) \right|^2 d\omega} \tag{6}$$

where the Fourier Transforms (FT) of x^{n+1}, $x(t)$, x_i, and λ are represented with X^{n+1}, $X(\omega)$, $X_i(\omega)$, and $\lambda(\omega)$, respectively. The real component of the inverse FT of $X^{n+1}(\omega)$ as filtered analytic signal comes from the mode in the time domain [30]. This study computed VMD-1D and -2D with $m = 1 \ldots 5$, $\alpha = 2000$, $\omega = 0$, $\tau = 0$, DC = 0, and init = 1 [33].

D. Feature Extraction

In the feature extraction step, texture-based measures were computed from the decomposition outputs using GLCM and GLDM. The two feature sets yielded feature vectors that summarize the underlying signal patterns in a structured form.

1) Gray-Level Co-occurrence Matrix (GLCM)

GLCM was introduced by Haralick *et al*. [34] in 1973 with fourteen basic features that is able to capture the image texture including thickness, uniformity, and other information. The frequency of pairs of pixels with particular gray-level values occurring in a 2D image is measured at particular distances and directions. Then, GLCM constructs a co-occurrence matrix which depicts the combinations of previously obtained gray-level values. GLCMs are computed using pixel pairs separated by distance (d) in pixels at orientation in degrees (θ) [34]. We used θ = 0°, 45°, 90°, 135° with d = 1. There were four extracted GLCM texture features used in this study, which were energy ($GLCM_E$), homogeneity ($GLCM_{HOM}$), correlation ($GLCM_{COR}$), and contrast ($GLCM_{CON}$), as formulated in Eqs. (7)–(10), in sequence.

$$GLCM_E = \sum_{i,j} (P(i,j))^2 \tag{7}$$

$$GLCM_{HOM} = \sum_{i,j} \frac{P(i,j)}{1 + |i - j|} \tag{8}$$

$$GLCM_{COR} = \sum_{i,j} \frac{(i - \mu_i)(i - \mu_j) P(i,j)}{\sigma_i \sigma_j} \tag{9}$$

$$GLCM_{CON} = \sum_{i,j} |i - j|^2 P(i,j) \tag{10}$$

where (i, j) represents the *GLCM* value, μ_i and μ_j are the mean values, σ_i and σ_j are the standard deviations, all of the i rows and j columns, respectively [35–37].

2) Gray-Level Difference Matrix (GLDM)

GLDM presents the characterization of an image using the difference between its adjacent pixels. GLDM observes the fluctuation intensity by computing the absolute Gray-level difference between adjacent pixels along the horizontal, vertical, and diagonal directions, as stated in Ref. [38]. Thus, in this study, the features of the image were calculated by implementing gradient contrast ($GLDM_{CON}$), gradient second moment ($GLDM_{SCM}$), gradient entropy ($GLDM_{ENT}$), gradient mean ($GLDM_{MEAN}$), and inverse-difference moment ($GLDM_{IDM}$). The equations of each feature are written in Eqs. (11)–(15).

$$GLDM_{CON} = \sum_g g^2 h(g|\theta) \tag{11}$$

$$GLDM_{SCM} = \sum_g \left[h(g|\theta) \right]^2 \tag{12}$$

$$GLDM_{ENT} = -\sum_g h(g|\theta) \log\left(h(g|\theta) \right) \tag{13}$$

$$GLDM_{MEAN} = \sum_g h(g|\theta)\, g \tag{14}$$

$$GLDM_{IDM} = \sum_g \frac{h(g|\theta)}{\left(g^2 + 1\right)} \tag{15}$$

where $h(g|\theta)$ is the probability of the amount of nearest-neighbor pixels with g gives the absolute delta value between two pixels in θ as its direction. In this

study, the direction degrees used in GLDM were the same as GLCM's property.

E. Classification

The extracted feature vectors were then provided to a supervised learning model to determine the corresponding lung sound category. In this work, an SVM classifier was used to learn the decision boundary from the training features and generate the final predictions on unseen samples.

1) Support Vector Machine (SVM)

SVM, as one of the ML classifiers introduced by Cortes and Vapnik [39], divides classes in the data by linearly separating the data using the hyperplane. The data closest to the hyperplane are termed Support Vectors (SV). The distance between the separating hyperplane and these SV defines the margin. A multi-class SVM projects samples into a high-dimensional feature space and identifies an optimal separating hyperplane by maximizing the margin between classes.

In some cases, the samples in the dataset cannot be classified by the hyperplane because they are linearly inseparable. So, SVM adopts a kernel function that converts the training data to a higher dimension through nonlinear mapping. There are several kernel functions, namely linear, quadratic, cubic, and Radial Basis Functions (RBF), which are calculated inside the nonlinear SVM using Eq. (16). Our study applied the RBF kernel and automated kernel scaling with the mathematical expression written in Eq. (17) to classify five classes that are listed in the lung sound dataset. Furthermore, the SVM was applied with no nested Cross-Validation (CV), no inner loop selecting gamma, and C was fixed at 2.0.

$$f(z)=\text{sign}\left(\sum_{i=1}^{SV}\alpha_i y_i K\left(z_i,z\right)+b\right) \tag{16}$$

$$K\left(z_i,z\right)=e^{-\frac{\left\|z_i^2-z^2\right\|}{2\sigma^2}} \tag{17}$$

where z_i and z are the feature vectors of training sample i and test sample, respectively, SV is the number of SV, and b is the bias value [39, 40].

F. Evaluation Metrics

The evaluation employed stratified 5-fold CV to ensure an unbiased estimate of generalization performance. Every fold used approximately 80% of the data for training and 20% for testing, with fold sizes differing by at most one sample due to the total 99 recordings in the lung sound dataset. To improve robustness and reduce variance from random partitioning, the 5-fold procedure was repeated 10 times with different random seeds. The accuracy, sensitivity, precision, and F1-Score were averaged across all folds and repetitions, which were derived using the following Eqs. (18)–(21) [41].

$$acc=\frac{TP+TN}{TP+FP+TN+FN} \tag{18}$$

$$prec=\frac{TP}{TP+FP} \tag{19}$$

$$sens=\frac{TP}{TP+FN} \tag{20}$$

$$F1-s=2\times\frac{prec\times sens}{prec+sens} \tag{21}$$

where acc is accuracy, $prec$ is precision, $sens$ is sensitivity, $F1-s$ is F1-Score, TP is true positive, FP is false positive, TN is true negative, and FN is false negative.

IV. Result and Discussion

The proposed STVMD-based model with GLCM and GLDM features for lung sound classification proposed in this study was modelled with MATLAB version R2023b and system specifications of Windows 11, AMD Ryzen 7 processor, and 16 GB RAM.

A. STVMD Results

The inputs for the STVMD stage were the pre-processed lung sound signals as displayed in Fig. 2 for each class's sample data. The flow then branches into two Strategies (STGs), which are STG 1 and STG 2. Following STG 1, TFR is decomposed with VMD-2D. Along with STG 2, the VMD-decomposed signals were proceeded to STFT. Both STGs led to five modes in the decomposed spectrogram per recording. Figs. 3 and 4 depict the decomposed signal results of both STGs from one sample recording per class.

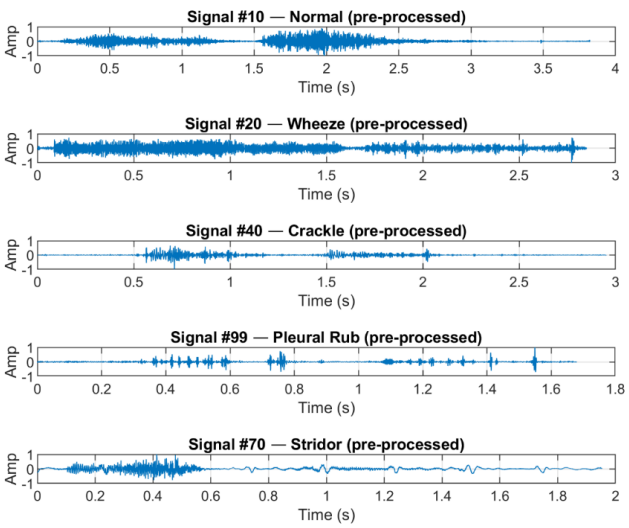


Fig. 2. Plot of the representative pre-processed signal of each lung sound class.

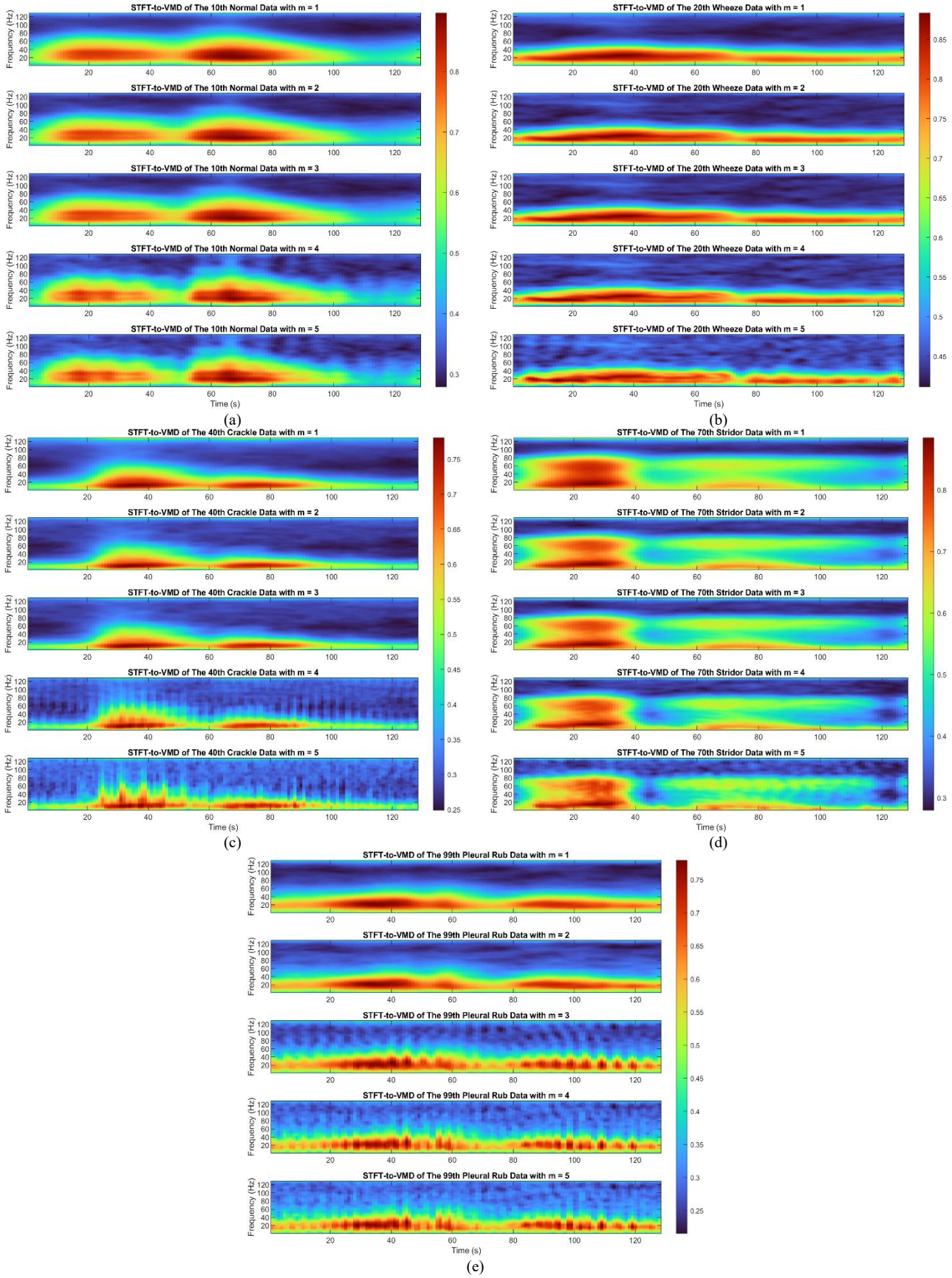


Fig. 3. Decomposed TFR result with STG 1 of each class: (a) normal, (b) wheeze, (c) crackle, (d) stridor, and (e) pleural rub, respectively.

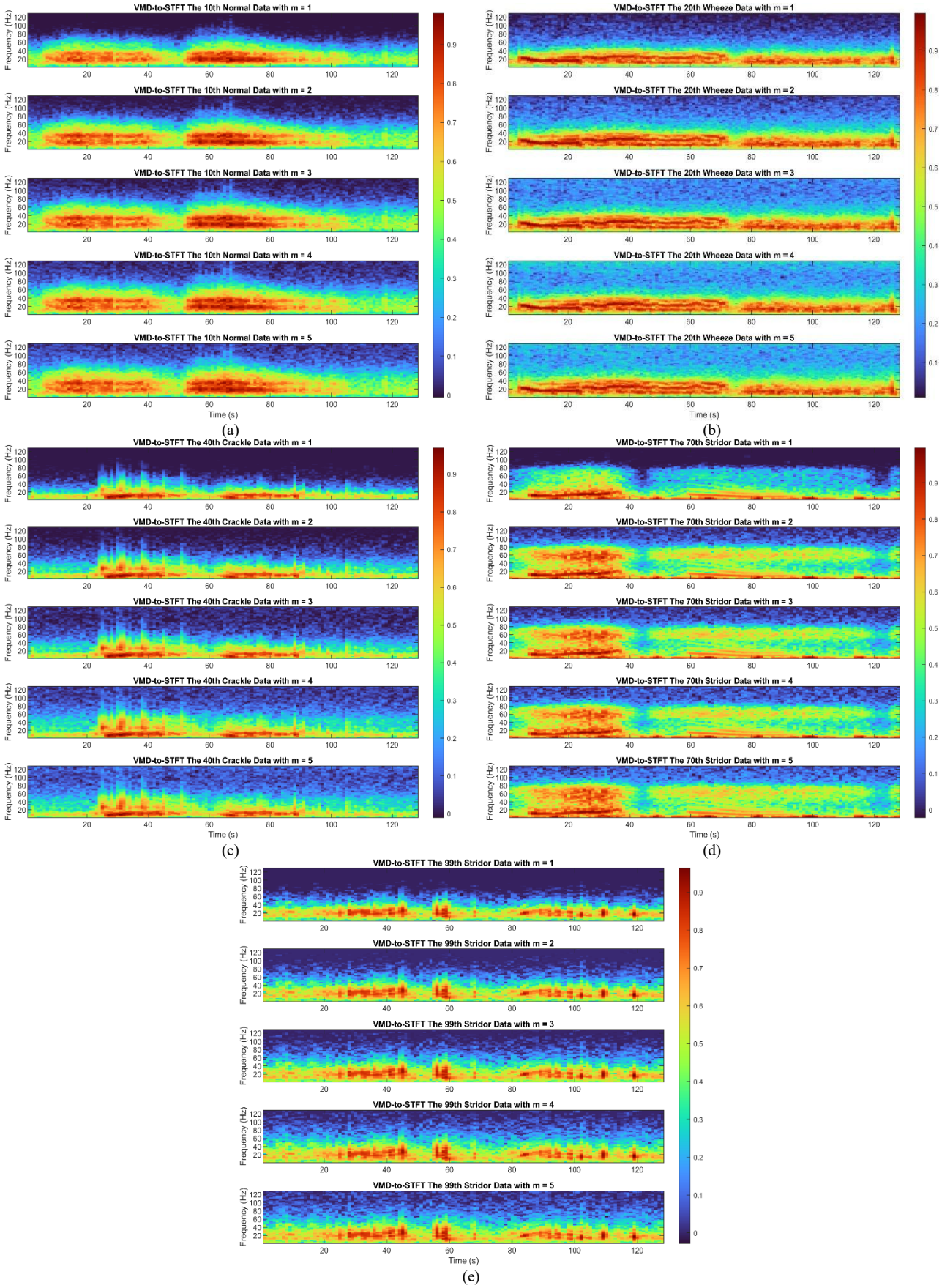


Fig. 4. Decomposed TFR result with STG 2 of each class: (a) normal, (b) wheeze, (c) crackle, (d) stridor, and (e) pleural rub, respectively.

B. Feature Extraction (FE) Results

In the Feature Extraction (FE) stage, our study used the arithmetic mean to consolidate direction-dependent texture descriptors into rotation-robust summary measures. As aforesaid, the GLCM and GLDM features are computed from four directions, yielding one value per direction. To minimize the total number of feature values and reduce sensitivity to preferential orientation in the image, the mean was counted and produced a single representative estimate per property.

Hence, before the mean calculation was applied, both methods gave 144 features. The details are that 16 and 20 features were given by GLCM and GLDM, respectively, after each feature was multiplied with 4 combinations of the STFT parameters. Then, after applying the mean formula, the total features dropped to 52, summed up from 32 and 20 features from GLCM and GLDM, orderly.

C. Classification Results and Analysis

To analyse the improvement of implementing VMD, Table II presents the baseline classification metric results obtained without deploying VMD. Then, the STVMD classification results are shown in Table III and Table IV for both STG 1 and STG 2. We classified the features from each mode one by one.

As detailed in Tables III and IV, the highest accuracy result was achieved in STG 1 using m = 1, with 97.98%, along with 98.04% of precision, 98.10% of sensitivity, and 98.02% of F1-Score. However, STG 1 is highly sensitive to the number of modes, where larger m values lead to substantial performance degradation and rapidly increasing computation time. This behavior indicates over-decomposition and reduced texture consistency for GLCM-GLDM features. In contrast, the STG 2 results using m = 2 produced slightly different accuracy at 97.37%, supported by precision, sensitivity, and F1-Score of 97.47%, 97.49%, and 97.43%, respectively. STG 2 yields consistently high performance for m = 2, 3, 4, suggesting greater durability to m. Regarding computational cost, STG 1 required up to 20.34 min to extract all features at m = 5, while STG 2 stayed below 8.6 min even at m = 5.

In addition, Table III summarizes the per-class precision and sensitivity values for all classes under STG 1 and STG 2. From Table III, it can be seen that STG 2 provided more consistent class-wise sensitivity across higher mode counts. STG 2 maintained 100% sensitivity for Normal and Pleural Rub for all modes, while STG 1 showed notable sensitivity degradation for these classes at larger m. Besides, STG 1 could be advantageous at a lower number of VMD modes, m = 1, by producing sharper class-specific texture separability, such as improved Wheeze sensitivity at 5.6% and higher Crackle precision at 4.8% compared to STG 2. However, STG 1 is more sensitive to mode selection, with substantial drops for the precision of Crackle at −30.8% and the sensitivity of Pleural Rub at −27.8%, both using m = 2 compared to STG 2. The lower percentages indicate the increased confusion between these acoustically similar classes under that setting. Taken together, STG 1 is best for peak performance at small m, whereas STG 2 better preserves stable per-class detection as m increases.

Fig. 5 presents the confusion matrices for the optimal configurations of both STGs. The values appearing in the diagonal matrix indicate that most of the samples were correctly classified among the five classes. Only a small number of misclassifications occurred in each strategy; for example, a single Crackle sample was misclassified as Pleural Rub, and one Stridor sample was misclassified as Normal in both scenarios.

TABLE II. CLASSIFICATION PERFORMANCE OF STG 1 AND STG 2 ACROSS VMD MODES (M) COMPARED TO THE NO-VMD BASELINE

VMD Mode (m)	Strategy	Accuracy (%)	Precision (%)	Sensitivity (%)	F1-Score (%)	Time for FE (min)
0	No VMD Baseline	96.46	96.57	96.60	96.54	0.18
1	**STG 1**	**97.98**	**98.04**	**98.10**	**98.02**	**0.16**
2	STG 1	90.91	91.81	90.78	91.12	1.25
3	STG 1	94.85	94.96	94.75	94.78	5.47
4	STG 1	89.09	89.69	89.17	89.27	9.82
5	STG 1	91.82	92.28	91.89	91.94	20.34
1	STG 2	96.36	96.42	96.41	96.37	1.37
2	STG 2	97.37	97.47	97.49	97.43	1.57
3	STG 2	97.17	97.26	97.32	97.25	5.17
4	**STG 2**	**97.27**	**97.36**	**97.43**	**97.35**	**7.31**
5	STG 2	95.86	96.06	96.06	95.98	8.51

TABLE III. PER-CLASS SENSITIVITY AND PRECISION RESULTS

STG	m	Normal (sens/prec)	Wheeze (sens/prec)	Crackle (sens/prec)	Pleural Rub (sens/prec)	Stridor (sens/prec)
STG 1 (%)	1	100/95.5	100/100	95.2/100	100/94.7	95.2/100
	2	95.2/100	94.4/94.4	85.7/69.2	72.2/86.7	95.2/100
	3	100/100	94.4/94.4	90.5/95	94.4/94.4	100/95.5
	4	85.7/94.7	94.4/94.4	90.5/86.4	83.3/88.2	95.2/87
	5	85.7/94.7	94.4/100	95.2/90.9	94.4/89.5	95.2/30.9
STG 2 (%)	1	100/95.5	94.4/100	95.2/95.2	100/94.7	95.2/100
	2	100/95.5	100/100	95.2/100	100/94.7	95.2/100
	3	100/95.5	100/100	95.2/100	100/94.7	95.2/100
	4	100/95.5	100/100	95.2/100	100/94.7	95.2/100
	5	100/95.5	100/100	90.5/100	100/94.7	95.2/95.2

TABLE IV. ABLATION RESULTS FOR STG 1 AND STG 2

STG	α	τ	Accuracy (%)	Precision (%)	Sensitivity (%)	F1-Score (%)	Time for FE (min)
STG 1	500	0	96.06	96.06	96.08	96.02	0.20
	500	0.1	96.46	96.59	96.56	96.51	11.47
	1000	0	96.67	96.73	96.81	96.74	0.21
	1000	0.1	96.57	96.67	96.67	96.61	13.55
	2000	0	97.98	98.04	98.10	98.02	0.16
	2000	0.1	96.06	96.20	96.17	96.13	12.37
	5000	0	90.10	92.27	90.03	90.31	0.24
	5000	0.1	94.95	95.10	95.05	95.03	13.01
	10,000	0	81.01	87.67	80.57	81.86	0.24
	10,000	0.1	94.85	95.10	94.92	94.96	15.26
STG 2	500	0	93.74	94.11	93.65	93.78	1.82
	500	0.1	95.96	96.01	96.11	95.99	2.96
	1000	0	95.25	95.30	95.37	95.29	1.73
	1000	0.1	94.65	95.03	94.52	94.66	2.87
	2000	0	97.37	97.47	97.49	97.43	1.57
	2000	0.1	95.96	96.13	96.05	96.04	4.43
	5000	0	95.96	96.01	96.11	95.99	2.03
	5000	0.1	97.47	97.56	97.62	97.54	4.67
	10,000	0	96.36	96.51	96.41	96.40	1.87
	10,000	0.1	93.74	93.95	93.79	93.72	9.97

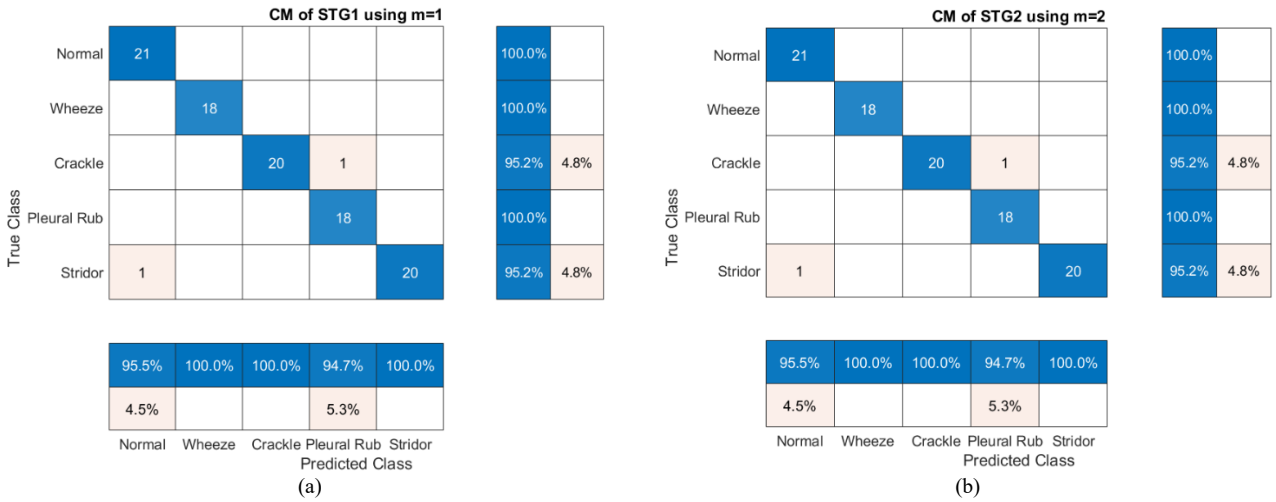


Fig. 5. Confusion matrix: (a) STG 1 using $m = 1$, (b) STG 2 using $m = 2$.

D. STVMD Analysis

From Table II, it is clearly explained that the VMD contributed to strengthening class divisibility by improving 1.52% and 0.91% of accuracy for STG 1 and STG 2, correspondingly. The qualitative samples from STVMD results of each class, as shown in Figs. 3 and 4, help explain why the two processing STGs behaved differently in the quantitative results presented in Table II. In STG 1, Fig. 3 presents the resulting 2D TFR maps across the 1st through the 5th modes, which appear visually similar with high smoothness. Computing VMD modes on STFT results produces enhanced TFR maps in the morphology features, which are harmonic ridges and structured textures. This makes GLCM and GLDM theoretically appropriate as they quantify spatial intensity relationships in the spectrogram.

In contrast, Fig. 4 depicts the TFRs of STG 2, which exhibit richer, more granular patterns across all deployed modes. Decomposing in the time domain first tends to yield modes with clearer band-limited behavior. This process signified that deeper 1D band-limited separation was beneficial when the signal is decomposed into finer components. This benefit comes at a substantially lower computational cost than STG 1 in higher m.

Overall, the qualitative observations and quantitative metrics together indicate that the processing order materially affected both interpretability and class separability. For this dataset and configuration, STG 1 is preferable for efficiency and peak accuracy at a low number of VMD modes. On the contrary, STG 2 can provide more persistent accuracy at higher modes where STG 1 tends to suffer from over-decomposition and texture fragmentation.

E. FE Analysis

Following the STG 1 result, the components were smooth and highly alike. Given that GLCM and GLDM are texture descriptors, the components reduced local contrast and co-occurrence variability. Continuing the process, the classifier was more sensitive to the selected setting and less robust.

Compared to the outputs from the STG 2, they were more favorable for GLCM-GLDM feature extraction methods. There were co-occurrence and intensity-difference statistics that maintained more

informative representation from the more structured textures rather than a smoothed envelope.

The FE runtime trends further validated this interpretation. As detailed in Table II, the FE time of STG 1 increased sharply as the decomposition setting increased, from 0.16 to 20.34 min, without a corresponding improvement in accuracy. This phenomenon indicates diminishing returns when VMD was applied after STFT. Contrarily, using STG 2, when VMD was employed before STFT, the accuracy values obtained were higher and more stable with substantially lower runtime, from 1.37 to 8.51 min at comparable settings. STG 2 offers a more balanced performance between discriminative power and computational cost for the lightweight GLCM- and GLDM-based SVM framework.

F. Ablation Study on VMD Parameters

In order to validate the robustness of the proposed STVMD framework and justify the selected VMD parameters, ablation experiments were conducted on the bandwidth penalty α and the noise tolerance τ [42–44]. The two parameters were computed at the optimal mode setting of each strategy, $m = 1$ for STG 1 and $m = 2$ for STG 2. In α, five values were tested, which were 500, 1000, 2000, 5000, and 10,000. As for τ, values of 0 and 0.1 were evaluated at each α level. Results are summarized in Table IV.

The ablation results reveal distinct sensitivity profiles between the two STGs. For STG 1, accuracy across all α and τ configurations spans 16.97 percentage points, peaking at $\alpha = 2000$, $\tau = 0$ with 97.98% and degrading sharply at higher α values. This sensitivity reflects the 2D nature of VMD operating on spectrograms, in that the excessively large α over-constrains the variational problem and disrupts the spatial texture structure required for discriminative GLCM-GLDM features. In contrast, STG 2 shows a substantially narrower accuracy range of 3.73 percentage points, confirming greater parametric robustness attributable to the more stable 1D decomposition on temporal waveforms.

Regarding τ, introducing a non-zero noise tolerance at $\tau = 0.1$ consistently increased feature extraction time by 11–15 min for STG 1 and 1–8 minutes for STG 2. Furthermore, the results came along with a negative accuracy benefit at the optimal α. Since the dataset is free of noise contamination, $\tau = 0$ is appropriate and yields substantially lower computational overhead. Overall, these experiments confirm that the originally selected parameters at $\alpha = 2000$, $\tau = 0$ represent the optimal configuration for both STGs.

G. Comparative Results Analysis

The comparative performance of this work's DSP-AI-based lung sound signal classification with various recent works is shown in Table V. A fundamental limitation in lung-sound research is the lack of standardized recording and pre-processing procedures, which can introduce inconsistencies across datasets and hinder direct performance comparisons. Therefore, the studies listed in Table V generally involved small cohorts and report different evaluation metrics, further limiting strict directly comparable benchmarking. Despite these constraints, the proposed approach demonstrates consistently high performance. Prior studies typically address fewer classes with accuracies ranging from approximately 91% to 97.3%, whereas our method achieved the accuracy-sensitivity values through STG 1 and STG 2, which are 97.98%–98.10% and 97.37%–97.43%, respectively. The close performance between STG 1 and STG 2 indicates stable classification behavior under different decomposition orderings, supporting the robustness of the proposed feature representation.

TABLE V. COMPARATIVE PERFORMANCE RESULTS WITH RECENT LUNG SOUND CLASSIFICATION METHODS

Authors	Year	Dataset	DSP Methods	Number of Classified Classes	Accuracy (%)	Sensitivity (%)
Kim *et al*. [20]	2022	Private dataset	Mel-spectogram	2	91.2	81
Rizal *et al*. [21]	2022	RALE® + 3M + NAU + UCLA 2015 data compilation	STFT + XSTFT / XMAX	2	100	Not specified
Dubey *et al*. [22]	2023	RALE®	WBS and WBP	3	95.69	95.69
Orkweha *et al*. [23]	2024	KAUH	MFCC	3	93.08	Not specified
				4	91.49	91.22
Engin *et al*. [24]	2026	Private dataset	Gammatonegram	4	97.3	97.75
Proposed study	2026	Kaggle 5-Class Lung Sounds Dataset	Proposed STVMD + GLCM + GLDM (STG 1 / STG 2)	5	97.98/97.37	98.10/97.43

Despite the competitive performance results reported, this research retains two prominent limitations that should be acknowledged. First, the small dataset that was used for the method evaluation contains the statistical robustness of the reported metrics. Not only that, but it also limits generalizability across diverse patient populations, recording devices, and clinical environments. Second, the FE time for STG 1 increased significantly when counted using a higher number of VMD modes, reaching over 20 min. This constrains real-time applicability in resource-limited clinical settings. One practical avenue to reduce computational overhead is to lower the STFT frequency resolution which directly reduces the spatial dimension of the input spectrogram and, consequently, the GLCM and GLDM computation cost. Nevertheless, given the relatively small dataset used

for evaluation, these findings should be interpreted as promising evidence of effectiveness and warrant further validation on larger and more diverse datasets.

V. Conclusion

This study addressed the need for DSP- and ML-based automated early-stage screening of lung diseases by proposing an STVMD-driven lung sound classification framework. As opposed to existing works where the TFR is typically used as the final input to learning models, the proposed approach explicitly decomposed the representation VMD components to build more stable and discriminative representations. Two alternative STGs were investigated: STG 1, which decomposed the STFT spectrogram using VMD-2D, and STG 2, which performed VMD-1D in the time domain before STFT. From each STVMD representation, features were obtained using GLCM and GLDM, and a multi-class SVM performed the final classification across five lung sound categories.

Experimental results on the five class Kaggle dataset showed that the proposed approach improves over the non-decomposed baseline, achieving the best overall performance of 97.98% accuracy using STG 1 and 97.37% using STG 2 under repeated stratified 5-fold CV. While STG 1 produced the highest accuracy, it also exhibited higher feature extraction time for larger mode settings, whereas STG 2 delivered more consistent performance with substantially lower runtime, making it more attractive for lightweight screening deployments. Future work will focus on extending validation across larger and more diverse patient cohorts and devices to support AI-on-device development.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

A.H. developed the concept for the paper, conducted the research work, analyzed the results, and drafted the manuscripts; A.R. prepared the dataset, investigated the research, and provided essential feedback for the manuscript writing; S.A. helped to visualize and validated the results; R.A.S. contributed to the abstract and conclusion sections of the manuscript; all authors had approved the final version.

References

[1] W. K. C. Morgan and A. Seaton, *Occupational Lung Diseases*. London, U.K.: W.B. Saunders Company, 1975.

[2] A. Ullah, M. S. Khan, M. U. Khan, and F. Mujahid, "Automatic classification of lung sounds using machine learning algorithms," in *Proc. Int. Conf. Frontiers Inf. Technol. (FIT)*, 2021, pp. 131–136. doi: 10.1109/FIT53504.2021.00033

[3] World Health Organization (WHO). (2026). Asthma. *World Health Organization*. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/asthma

[4] Z. Mao *et al*., "Global, regional, and national burden of asthma from 1990 to 2021: A systematic analysis of the global burden of disease study 2021," *Chin. Med. J. Pulm. Crit. Care Med.*, vol. 3, no. 1, pp. 50–59, 2025. doi: 10.1016/j.pccm.2025.02.005

[5] World Health Organization (WHO). (2026). Chronic obstructive pulmonary disease (COPD). *World Health Organization*. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-%28copd%29

[6] S. Safiri *et al*., "Global burden of lower respiratory infections during the last three decades," *Front. Public Health*, vol. 10, 1028525, 2022. doi: 10.3389/fpubh.2022.1028525

[7] N. Adderley, J. Goldin, and S. Sharma, *Pleural Friction Rub*. Treasure Island, FL, USA: StatPearls Publishing, 2024.

[8] A. Bohadana, G. Izbicki, and S. S. Kraman, "Fundamentals of lung auscultation," *N. Engl. J. Med.*, vol. 370, no. 8, pp. 744–751, 2014. doi: 10.1056/nejmra1302901

[9] M. Aykanat, Ö. Kılıç, B. Kurt, and S. Saryal, "Classification of lung sounds using convolutional neural networks," *EURASIP J. Image Video Process.*, vol. 2017, no. 1, 65, 2017. doi: 10.1186/s13640-017-0213-2

[10] X. Xu and R. Sankar, "Classification and recognition of lung sounds using artificial intelligence and machine learning: A literature review," *Big Data Cogn. Comput.*, vol. 8, no. 10, 127, 2024. doi: 10.3390/bdcc8100127

[11] J. P. Garcia-Mendez *et al*., "Machine learning for automated classification of abnormal lung sounds obtained from public databases: A systematic review," *Bioengineering*, vol. 10, no. 10, 1155, 2023. doi: 10.3390/bioengineering10101155

[12] Z. Neili and K. Sundaraj, "A comparative study of the spectrogram, scalogram, melspectrogram and gammatonegram time-frequency representations for the classification of lung sounds using the ICBHI database based on CNNs," *Biomed. Tech.*, vol. 67, no. 5, pp. 367–390, 2022. doi: 10.1515/bmt-2022-0180

[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, 586, 2024. doi: 10.3390/bioengineering11060586

[14] B. Acar Demirci, Y. Koçyiğit, D. Kizilirmak, and Y. Havlucu, "Adventitious and normal respiratory sound analysis with machine learning methods," *Celal Bayar Univ. J. Sci.*, vol. 18, no. 2, pp. 169–180, 2021. doi: 10.18466/cbayarfbe.1002917

[15] A. Rizal and A. Puspitasari, "Lung sound classification using wavelet transform and entropy to detect lung abnormality," *Serbian J. Electr. Eng.*, vol. 19, no. 1, pp. 79–98, 2022. doi: 10.2298/SJEE2201079R

[16] J. Amose *et al.*, "Classification of adventitious lung sounds: Wheeze, crackle using machine learning techniques," *Int. J. Intell. Syst. Appl. Eng.*, vol. 11, no. 3, pp. 1143–1152, 2023.

[17] B. Hou, D. Wang, T. Xia, Z. Peng, and K.-L. Tsui, "Difference mode decomposition for adaptive signal decomposition," *Mech. Syst. Signal Process.*, vol. 191, 110203, 2023. doi: 10.1016/j.ymssp.2023.110203

[18] Y. Liu *et al*., "An efficient new adaptive variational mode decomposition algorithm for extracting adventitious lung sounds," *Biomed. Signal Process. Control*, vol. 89, 105700, 2024. doi: 10.1016/j.bspc.2023.105700

[19] Y. Ma, X. Xu, and Y. Li, "LungRN+NL: An improved adventitious lung sound classification using non-local block ResNet neural network with mixup data augmentation," in *Proc. Interspeech*, 2020, pp. 2902–2906. doi: 10.21437/Interspeech.2020-2487

[20] B. J. Kim, B. S. Kim, J. H. Mun, C. Lim, and K. Kim, "Author correction: An accurate deep learning model for wheezing in children using real world data," *Sci. Rep.*, vol. 13, no. 1, 1156, 2023. doi: 10.1038/s41598-023-28376-8

[21] A. Rizal, W. Priharti, D. Rahmawati, and H. Mukhtar, "Classification of pulmonary crackle and normal lung sound using spectrogram and support vector machine," *J. Biomim. Biomater. Biomed. Eng.*, vol. 55, pp. 143–153, 2022. doi: 10.4028/p-tf63b7

[22] R. Dubey, R. M. Bodade, and D. Dubey, "Efficient classification of the adventitious sounds of the lung through a combination of SVM-LSTM-Bayesian optimization algorithm with features based on wavelet bi-phase and bi-spectrum," *Res. Biomed. Eng.*, vol. 39, no. 2, pp. 349–363, 2023. doi: 10.1007/s42600-023-00270-2

[23] K. Orkweha *et al*., "A framework for detecting pulmonary diseases from lung sound signals using a hybrid multi-task autoencoder-SVM model," *Symmetry*, vol. 16, no. 11, 1413, 2024. doi: 10.3390/sym16111413

[24] M. A. Engin, R. Uzun Arslan, İ. Senyer Yapici, S. Aras, and A. Gangal, "Deep learning-based classification of common lung

sounds via auto-detected respiratory cycles," *Bioengineering*, vol. 13, no. 2, p. 170, 2026. doi: 10.3390/bioengineering13020170
[25] Kaggle. (2026). 5-Class Lung Sounds Dataset from Several Sources. [Online]. Available: https://www.kaggle.com/datasets/annisahmrn/5-class-lung-sounds-dataset-from-several-sources
[26] R.A.L.E. Repository. (2025). The R.A.L.E repository. [Online]. Available: http://www.rale.ca/Repository.htm
[27] A. Rizal, R. Hidayat, H. A. Nugroho, and W. A. Cahyadi, "Lung sound classification using multiresolution Higuchi fractal dimension measurement," *Int. J. Electr. Comput. Eng.*, vol. 13, no. 5, pp. 5091–5100, 2023. doi: 10.11591/ijece.v13i5.pp5091-5100
[28] A. M. Mekala and S. Chandrasekaran, "Heart sound interference cancellation from lung sound using dynamic neighbourhood learning-particle swarm optimiser based optimal recursive least square algorithm," *Int. J. Biomed. Eng. Technol.*, vol. 34, no. 2, 133, 2020. doi: 10.1504/IJBET.2020.111000
[29] A. Rizal, R. Hidayat, and H. A. Nugroho, "Pulmonary crackle feature extraction using tsallis entropy for automatic lung sound classification," in *Proc. 1st Int. Conf. Biomed. Eng. (IBIOMED)*, 2016, pp. 1–4. doi: 10.1109/IBIOMED.2016.7869823
[30] K. Keerthi Krishnan and K. P. Soman, "CNN based classification of motor imaginary using variational mode decomposed EEG-spectrum image," *Biomed. Eng. Lett.*, vol. 11, no. 3, pp. 235–247, 2021. doi: 10.1007/s13534-021-00190-z
[31] A. Rizal, R. Hidayat, and H. A. Nugroho, "Lung sounds classification using spectrogram's first order statistics features," in *Proc. 6th Int. Annu. Eng. Seminar (InAES)*, 2016, pp. 96–100. doi: 10.1109/INAES.2016.7821914
[32] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, 2014. doi: 10.1109/TSP.2013.2288675
[33] D. Liu, B. Awudong, X. Li, and Q. Li, "Research on motor imaging EEG signals based on VMD," in *Proc. 4th Int. Conf. Bioinf. Intell. Comput.*, 2024, pp. 147–152. doi: 10.1145/3665689.3665714
[34] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-3, no. 6, pp. 610–621, 1973. doi: 10.1109/TSMC.1973.4309314
[35] A. F. H. Alharan, H. K. Fatlawi, and N. S. Ali, "A cluster-based feature selection method for image texture classification," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 14, no. 3, 1433, 2019. doi: 10.11591/ijeecs.v14.i3.pp1433-1442
[36] Istiqomah, R. Y. D. Wardhana, and A. Rizal, "Classification of heart sounds using grey level co-occurrence matrix and logistic regression," *Math. Model. Eng. Probl.*, vol. 11, no. 7, pp. 1899–1905, 2024. doi: 10.18280/mmep.110720
[37] K. A. Zahroh, D. C. R. Novitasari, and L. Hakim, "Perbandingan ekstraksi fitur untuk klasifikasi COVID-19, MERS, dan SARS menggunakan algoritma extreme learning machine," *J. Fourier*, vol. 13, no. 1, pp. 30–41, 2024. doi: 10.14421/fourier.2024.131.30-41
[38] J. S. Weszka, C. R. Dyer, and A. Rosenfeld, "A comparative study of texture measures for terrain classification," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-6, no. 4, pp. 269–285, 1976. doi: 10.1109/TSMC.1976.5408777
[39] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995. doi: 10.1007/BF00994018
[40] K. Devi Thangavel, U. Seerengasamy, S. Palaniappan, and R. Sekar, "Prediction of factors for controlling of green house farming with fuzzy based multiclass support vector machine," *Alex. Eng. J.*, vol. 62, pp. 279–289, 2023. doi: 10.1016/j.aej.2022.07.016
[41] G. Amelia, A. Humairani, T. W. Purboyo, and D. Naufal, "ECG arrhythmia detection using complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN)," in *Proc. 9th Int. Conf. Instrum. Control Autom. (ICA)*, 2025, pp. 128–133. doi: 10.1109/ICA65945.2025.11251891
[42] Q. Huang, X. Liu, Q. Li, Y. Zhou, T. Yang, and M. Ran, "A parameter-optimized variational mode decomposition method using salp swarm algorithm and its application to acoustic-based detection for internal defects of arc magnets," *AIP Adv.*, vol. 11, no. 6, 2021. doi: 10.1063/5.0054894
[43] T. Eriksen and N. ur Rehman, "Data-driven nonstationary signal decomposition approaches: A comparative analysis," *Sci. Rep.*, vol. 13, no. 1, 1798, 2023. doi: 10.1038/s41598-023-28390-w
[44] O. Ahmad, L. Wesemann, F. Waschkowski, and Z. Khalid, "Variational mode-driven graph convolutional network for spatiotemporal traffic forecasting," arXiv Preprint, arXiv:2408.16191, 2025.

Copyright © 2026 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (CC BY 4.0).