

Article

Overcoming Data Scarcity in Non-Contact Respiratory Monitoring: A DTW-Enhanced CNN-LSTM Approach

Ju O Kim and Deokwoo Lee * 

Department of Computer Engineering, Keimyung University, Daegu 42601, Republic of Korea; 1113199@stu.kmu.ac.kr

* Correspondence: dwoolee@kmu.ac.kr; Tel.: +82-53-580-5234

Abstract

This study investigates non-contact respiratory pattern classification using Ultra-Wideband (UWB) radar sensors and deep learning. A CNN-LSTM hybrid architecture was developed combining spatial feature extraction through convolutional layers with temporal pattern recognition via LSTM networks. To address data scarcity in the minority class, a two-stage augmentation strategy incorporating Dynamic Time Warping-based SMOTE-TS was implemented. The experimental evaluation utilized 700 respiratory recordings from seven healthy volunteers performing controlled breathing exercises. Under controlled laboratory conditions, the system achieved 94.3% accuracy and 0.969 AUC, with an average inference time of 45.3 ms per sample (SD: 8.7 ms), demonstrating computational feasibility for real-time applications. This preliminary investigation establishes technical proof-of-concept, though validation with clinical populations remains necessary before medical deployment.

Keywords: CNN-LSTM; deep learning; respiratory monitoring; UWB radar; data augmentation; SMOTE-TS; non-contact sensing



Academic Editor: Michael Winter

Received: 16 September 2025

Revised: 17 October 2025

Accepted: 20 October 2025

Published: 21 October 2025

Citation: Kim, J.O.; Lee, D. Overcoming Data Scarcity in Non-Contact Respiratory Monitoring: A DTW-Enhanced CNN-LSTM Approach. *Electronics* **2025**, *14*, 4120. <https://doi.org/10.3390/electronics14204120>

Copyright: © 2025 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

Respiratory monitoring serves as a fundamental physiological indicator in healthcare settings [1]. Abnormal breathing patterns, including apnea, hypopnea, and irregular respiratory rhythms, correlate with various medical conditions such as sleep apnea syndrome, Chronic Obstructive Pulmonary Disease (COPD), and cardiac disorders [2]. Early detection of these abnormalities enables timely intervention and improved patient outcomes. Recent epidemiological studies indicate that sleep-disordered breathing affects 936 million adults aged 30–69 years globally [3].

Traditional respiratory monitoring methods require direct contact with patients through chest-mounted sensors or nasal cannulas [4]. While these approaches provide accurate measurements, they often cause discomfort, restrict movement, and may lead to skin irritation during extended monitoring periods. These limitations particularly affect vulnerable populations, including neonates, burn patients, and individuals with sensitive skin conditions. Non-contact respiratory monitoring has emerged as an alternative approach, enabling continuous measurement without physical sensors, potentially improving patient comfort and compliance [5]. Recent comprehensive surveys have highlighted the rapid evolution of machine learning approaches for radar-based vital sign monitoring, demonstrating significant advances in accuracy and reliability [6,7].

Among available non-contact technologies, Ultra-Wideband (UWB) radar demonstrates distinct advantages through high temporal resolution, clothing penetration capabil-

ity, and independence from lighting conditions [8]. UWB radar systems transmit ultra-short electromagnetic pulses and measure time-of-flight variations to detect minute chest wall movements associated with breathing. Recent advances in deep learning have transformed biosignal analysis, enabling automatic feature extraction and pattern recognition from complex physiological data [9]. Convolutional Neural Networks (CNNs) effectively extract spatial features from time-series data, while Long Short-Term Memory (LSTM) Networks capture temporal dependencies [10]. Hybrid architectures combining these approaches have demonstrated superior performance in various time-series classification tasks. Recent studies have shown promising results using transfer learning approaches for UWB radar-based heart rate monitoring, demonstrating the potential for knowledge transfer across different radar configurations [11].

However, applying deep learning to medical applications faces significant challenges, particularly the scarcity of labeled abnormal respiratory samples. Medical data collection encounters constraints from ethical considerations, patient availability, and the rarity of certain conditions. This data imbalance can severely impact model training, leading to biased predictions favoring the majority class. Novel approaches, such as denoising diffusion probabilistic models, have recently been proposed to address signal quality challenges in UWB radar-based monitoring [12].

1.1. Scope and Limitations of This Study

This preliminary investigation focuses on establishing technical feasibility rather than providing a clinically validated system. The study utilizes simulated respiratory abnormalities from healthy volunteers in controlled laboratory conditions rather than clinical populations. The dataset comprises seven participants with 700 recordings in total, collected under controlled conditions with fixed sensor distance and minimal interference. While these constraints enable rigorous algorithm development, extensive clinical validation with diverse patient populations remains necessary before medical deployment.

1.2. Research Contributions

Within the constraints described above, this study makes the following technical contributions. First, a CNN-LSTM hybrid architecture is demonstrated for respiratory pattern classification in controlled settings, achieving 94.3% accuracy on simulated abnormality detection. Second, a two-stage augmentation approach combining basic time-series transformations with DTW-based SMOTE-TS is developed, demonstrating effectiveness for addressing class imbalance in limited respiratory datasets. Third, a baseline architecture for future work is provided, including complete implementation details that can serve as a foundation for subsequent clinical validation studies. Fourth, computational efficiency with real-time processing capability at 45.3 ms inference time is established, demonstrating feasibility for continuous monitoring applications.

The remainder of this paper is organized as follows: Section 2 reviews related work in non-contact respiratory monitoring and deep learning approaches; Section 3 describes the methodology, including data collection, preprocessing, augmentation, and model architecture; Section 4 presents experimental results and comparisons; Section 5 discusses technical achievements, limitations, and future directions; and Section 6 concludes with a summary of findings and implications for future research.

2. Related Work

2.1. Non-Contact Respiratory Monitoring Technologies

Non-contact respiratory monitoring technologies include optical imaging, thermal imaging, and radar-based approaches [13]. Optical methods detect breathing through subtle

color or movement changes but require adequate lighting and are sensitive to ambient conditions [14]. Thermal imaging measures temperature variations during respiration but faces challenges with clothing interference and requires relatively close proximity [15].

Radar-based approaches, particularly UWB and millimeter-wave (mmWave) radar, have gained attention for their robustness across various environmental conditions. Several recent studies have demonstrated contactless vital sign monitoring capabilities. Ahmad et al. [16] developed an mmWave efficient modulatorless tracking radar system for vital signs detection, demonstrating the feasibility of simplified radar architectures. For multi-person scenarios, Lee et al. [17] explored FMCW radar configurations for vital sign monitoring, though the present single-person study does not address this complexity. Leem et al. [18] investigated impulse-radio UWB radar for vital signs extraction using signal processing approaches, whereas the current work emphasizes deep learning methods for pattern classification. Recent work by Gruzewska et al. [11] has demonstrated the potential of transfer learning approaches for UWB radar-based monitoring, enabling knowledge transfer between different radar configurations.

2.2. Deep Learning for Time-Series Classification

Deep learning architectures have revolutionized time-series analysis in medical applications [19]. CNNs excel at extracting local patterns through convolutional filters, automatically learning relevant features from raw signals. RNNs and their variants, particularly LSTM networks, address the challenge of capturing temporal dependencies by maintaining internal states that encode historical context [10].

Hybrid CNN-LSTM architectures leverage the complementary strengths of both approaches: CNNs for spatial feature extraction and LSTMs for temporal sequence modeling. Recent studies demonstrate that such hybrid architectures outperform individual models in complex time-series tasks, including physiological signal analysis [20]. However, most prior work has access to larger datasets or clinical recordings, whereas the present study specifically addresses the challenge of limited data availability through augmentation strategies. Khan et al. [21] recently proposed hybrid deep learning models for UWB radar-based human activity recognition, demonstrating the versatility of these architectures beyond vital sign monitoring.

2.3. Data Augmentation in Medical Applications

Data augmentation addresses the scarcity of medical training samples through synthetic data generation [22]. Basic techniques for time-series include time shifting, amplitude scaling, and noise injection, which increase dataset diversity without changing fundamental signal characteristics.

Advanced methods leverage domain-specific properties. Dynamic Time Warping (DTW) provides a distance metric for temporal sequences that accounts for variations in speed or timing [23]. SMOTE-TS extends the Synthetic Minority Over-Sampling Technique to time-series data, generating synthetic samples by interpolating between nearest neighbors in the minority class [24]. The combination of DTW distance metric with SMOTE-TS interpolation has shown promise for addressing class imbalance in medical time-series [25]. Novel approaches using diffusion models have also emerged, with Wang et al. [12] demonstrating improved signal quality through denoising diffusion probabilistic models for UWB radar-based respiration monitoring.

This work combines DTW-based SMOTE-TS with traditional augmentation methods, specifically for the analysis of respiratory signals from UWB radar, an application context that has received limited attention in prior augmentation literature.

2.4. Comparative Summary of Related Work

Table 1 provides a systematic comparison of this study with related non-contact respiratory monitoring research. This comparison clarifies the positioning and contributions of the current work.

Table 1. Comparison with related non-contact respiratory monitoring systems.

Study	Sensor	Data Source	Method	Validation	Key Difference
[16]	mmWave radar	Laboratory	Modulator-less tracking	Laboratory	Different radar technology
[17]	FMCW radar	Multiple subjects	FMCW Doppler	Controlled	Multi-person focus
[18]	IR-UWB	Vital signs	Signal processing	Laboratory	No deep learning
[26]	UWB	Sleep apnea patients	CNN	Clinical data	Clinical validation
[27]	Multi-modal	Activity data	CNN-LSTM	Sensors	Different application
This work	UWB	Simulated ($n = 7$)	CNN-LSTM + DTW-SMOTE	Controlled lab	Focus on augmentation

Several key distinctions emerge from this comparison. Regarding dataset characteristics, unlike studies with clinical data [26] or large-scale datasets [27], this work explicitly addresses the challenge of extremely limited data through augmentation strategies. In terms of methodological focus, while prior work emphasizes sensor hardware [16,17] or signal processing [18], this study focuses on deep learning architecture and data augmentation for handling class imbalance. Concerning clinical validation status, studies using patient data [26] provide clinically relevant validation, whereas this preliminary study uses simulated abnormalities, clearly positioning it as a technical feasibility investigation. Finally, regarding augmentation approach, the combination of DTW-based SMOTE-TS with CNN-LSTM architecture represents a specific contribution to addressing minority class scarcity in respiratory time-series.

3. Materials and Methods

3.1. Experimental Data Collection

3.1.1. Participants and Protocol

Seven healthy volunteers (lab members) were recruited for this technical feasibility study (four males, three females; age: 25.4 ± 3.2 years; BMI: 22.8 ± 2.1 kg/m²). Each participant performed controlled breathing exercises, including normal breathing at natural respiration rates (12–20 breaths/min), simulated apnea through voluntary breath-holding for 10–15 s, and irregular breathing with intentional variations in breathing depth and frequency.

Data collection occurred in a controlled laboratory environment with participants seated approximately 1 m from the UWB radar sensor. Each recording session lasted 60 s, with multiple trials per participant. The complete dataset comprises 700 recordings: 500 classified as normal breathing and 200 as abnormal patterns. Figure 1 illustrates the complete system architecture from signal acquisition to classification output.

3.1.2. UWB Radar Configuration

UWB radar signals were acquired using a commercial sensor operating at appropriate frequency range with a sampling rate of 10 Hz. The radar was positioned to capture chest wall movements in the anterior–posterior direction. Signal acquisition parameters were selected to balance temporal resolution with computational requirements.

Professional System Architecture for UWB Radar-based Respiratory Pattern Detection

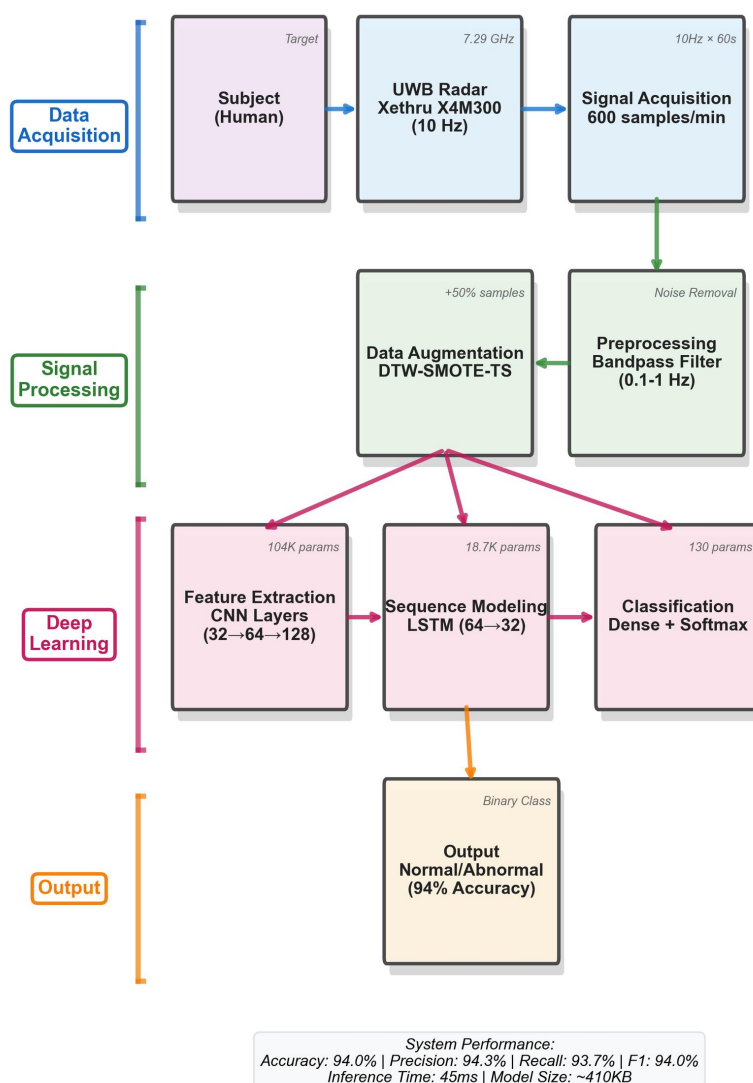


Figure 1. System architecture overview.

3.1.3. Dataset Composition and Splitting

The dataset was partitioned using stratified splitting to maintain class balance across subsets. The training set comprised 60% (420 samples), the validation set 20% (140 samples), and the test set 20% (140 samples). Stratified splitting ensures equal class proportions in each subset.

Due to the limited number of participants ($n = 7$), subject-level splitting was not feasible, as it would result in extremely small test sets that could not reliably evaluate model performance. Therefore, recording-level splitting was employed, which may lead to optimistic performance estimates, as discussed in Section 4.4. This represents a limitation of the current study and motivates the need for larger-scale validation, as discussed in Section 5.4.

3.2. Data Preprocessing

Raw UWB radar signals undergo preprocessing to enhance signal quality and remove artifacts. A fourth-order Butterworth bandpass filter (0.1–2 Hz) is applied to isolate respira-

tory frequencies while attenuating high-frequency noise and baseline drift. This frequency range encompasses typical human respiratory rates (6–120 breaths per minute).

Signal normalization employs per-window z-score standardization, as shown in Equation (1):

$$x_{norm} = \frac{x - \mu}{\sigma} \quad (1)$$

where x represents the raw signal, μ denotes the mean, and σ indicates the standard deviation calculated over each 60 s window. This normalization approach obscures inter-subject amplitude variability, which may carry diagnostic information. This design choice prioritizes model convergence in the limited dataset over preserving absolute amplitude information. Future work should investigate alternative normalization strategies that retain amplitude features. Representative examples of preprocessed respiratory patterns are shown in Figure 2.

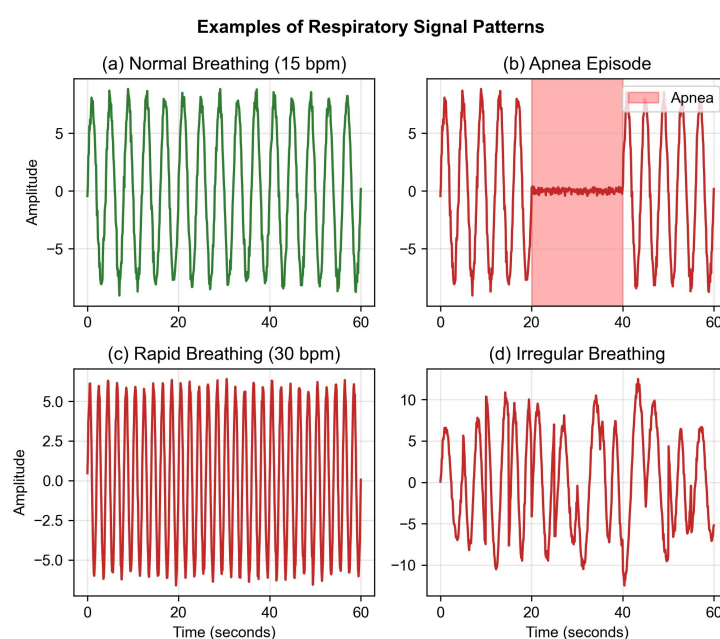


Figure 2. Representative respiratory patterns from UWB radar.

3.3. Data Augmentation Strategy

To address the class imbalance between normal (71.4%) and abnormal (28.6%) samples, a two-stage augmentation approach was implemented, applied only to the training set.

3.3.1. Basic Augmentation

Three fundamental time-series transformations were applied. Time shifting applies random circular shifts within $\pm 10\%$ of signal length to simulate temporal alignment variations. This range preserves breathing cycle structure while introducing sufficient diversity. Amplitude scaling multiplies signals by factors uniformly sampled from $[0.8, 1.2]$ to simulate variations in breathing depth or sensor distance. These bounds ensure physiologically plausible amplitude ranges based on typical chest wall displacement (5–15 mm). Gaussian noise injection adds white noise with $\text{SNR} = 20$ dB to simulate sensor noise and environmental interference. The SNR level was selected based on typical UWB radar noise characteristics observed in preliminary measurements.

3.3.2. DTW-Based SMOTE-TS

SMOTE-TS generates synthetic minority class samples through interpolation between nearest neighbors identified using DTW distance. DTW provides a similarity metric robust to temporal variations in breathing cycles, as defined in Equation (2):

$$d_{DTW}(X, Y) = \min_{\pi} \sqrt{\sum_{k=1}^K (x_{\pi_1(k)} - y_{\pi_2(k)})^2} \quad (2)$$

where π represents the optimal warping path aligning sequences X and Y , and K is the total number of aligned points. The DTW algorithm finds the warping path that minimizes the cumulative distance between sequences, making it particularly suitable for respiratory signals, where breathing cycles naturally vary in duration across individuals and conditions.

Synthetic samples are created through interpolation according to Equation (3):

$$S_{new} = S_i + \lambda \cdot (S_{nn} - S_i) \quad (3)$$

where S_i is a randomly selected seed sample from the minority class, S_{nn} is one of its k -nearest neighbors ($k = 5$) based on DTW distance, and $\lambda \sim U(0.2, 0.8)$ controls interpolation strength. The $k = 5$ parameter balances local structure preservation with diversity, while the bounds on λ prevent extreme interpolation that might create artifacts.

The rationale for DTW-SMOTE-TS in respiratory signal augmentation is threefold. First, respiratory patterns exhibit natural temporal variability in cycle duration, which DTW accommodates by flexible alignment. Second, interpolation in the DTW-aligned space preserves the quasi-periodic structure of breathing signals. Third, generating synthetic minority samples helps mitigate the class imbalance problem that would otherwise bias the model toward normal breathing patterns. Figure 3 demonstrates the various augmentation techniques applied to respiratory signals.

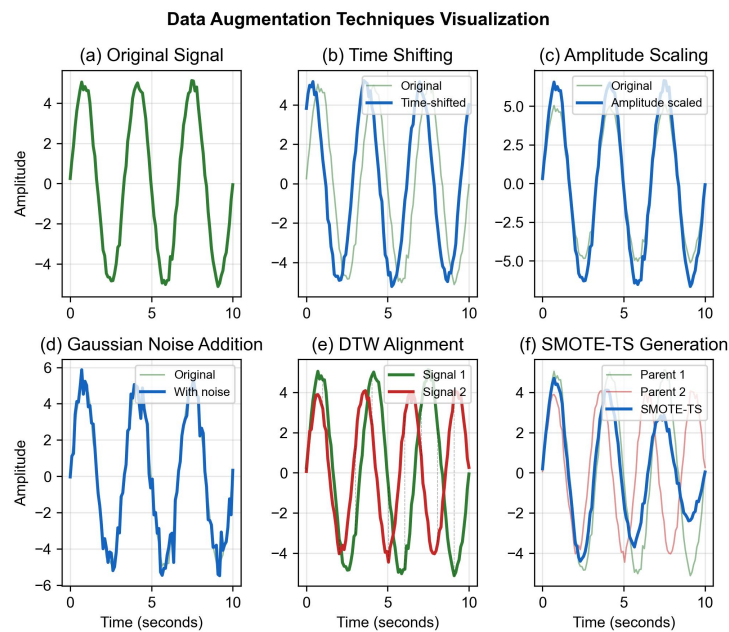


Figure 3. Data augmentation examples on respiratory signals.

3.4. CNN-LSTM Architecture

3.4.1. Architecture Design Rationale

The choice of CNN-LSTM hybrid architecture is motivated by the characteristics of respiratory signals and the requirements of the classification task. Convolutional layers extract local patterns from time-series data. Respiratory signals exhibit quasi-periodic structures (breathing cycles) with characteristic waveforms (inhale–exhale patterns). CNNs automatically learn relevant features from these patterns through hierarchical filter banks, eliminating the need for manual feature engineering. Progressively smaller kernels (15, 7, 3) capture features at multiple temporal scales.

LSTM layers model temporal dependencies between breathing cycles. Abnormal patterns often involve changes in cycle-to-cycle regularity, phase relationships, or long-term trends. LSTMs maintain internal memory states that encode historical context, enabling detection of such temporal patterns. Bidirectional processing captures both forward and backward temporal context.

The hybrid approach combines these strengths: CNNs extract robust local features invariant to small temporal shifts, while LSTMs integrate these features over time to recognize pattern sequences. This design is particularly effective for respiratory signals, where both local waveform shape and temporal regularity carry diagnostic information.

3.4.2. Network Architecture

The hybrid model processes 600-dimensional input vectors ($60\text{ s} \times 10\text{ Hz}$) through three convolutional blocks with progressively increasing filter counts and decreasing kernel sizes. Block 1: Conv1D (32 filters, kernel = 15) + BatchNorm + ReLU + MaxPool(2) + Dropout(0.3). Block 2: Conv1D (64 filters, kernel = 7) + BatchNorm + ReLU + MaxPool(2) + Dropout(0.3). Block 3: Conv1D (128 filters, kernel = 3) + BatchNorm + ReLU + GlobalMaxPool + Dropout(0.3).

The LSTM temporal modeling stage consists of two layers: LSTM Layer 1 (64 units, `return_sequences = True`) and LSTM Layer 2 (32 units, `return_sequences = False`). The classification stage comprises Dense Layer (64 units, ReLU, Dropout = 0.3) followed by Output Layer (2 units, Softmax).

The total model contains 287,842 trainable parameters. Batch normalization after each convolutional layer accelerates training and improves stability. Dropout ($p = 0.3$) provides regularization to prevent overfitting given the limited dataset size. The complete architecture is illustrated in Figure 4.

3.5. Training Configuration

Hyperparameters were determined through systematic grid search on the validation set. The optimal configuration was identified as learning rate $\alpha = 0.001$, batch size = 32, and dropout rate = 0.3. Higher learning rates (0.01) caused training instability with oscillating losses, while lower rates (0.0001) converged too slowly. Batch size 32 provided a good balance between gradient estimate quality, memory constraints, and training speed. Dropout rate 0.3 provided sufficient regularization without excessive information loss. For architectural parameters, kernel sizes (15, 7, 3) were chosen to capture multi-scale temporal features corresponding to complete breathing cycles, partial cycles, and fine-grained waveform details, respectively.

Model training employed the Adam optimizer with default momentum parameters ($\beta_1 = 0.9$, $\beta_2 = 0.999$) and categorical cross-entropy as the loss function. The maximum number of epochs was set to 100, with early stopping implemented with patience of 20 epochs monitoring validation loss. A learning rate schedule using ReduceLROnPlateau was applied with reduction factor 0.5 and patience 10 epochs.

All experiments were conducted using NVIDIA GTX 1080 Ti GPU (11 GB VRAM) and Intel Xeon E5-2690 CPU. The software environment consisted of Python 3.8, TensorFlow 2.10, CUDA 11.2, and cuDNN 8.1. Training time was approximately 45 min per fold in 5-fold cross-validation, with total training time of approximately 4 h for all experiments, including hyperparameter search. Peak GPU memory usage reached 8.2 GB.

CNN-LSTM Hybrid Architecture for Respiratory Pattern Classification

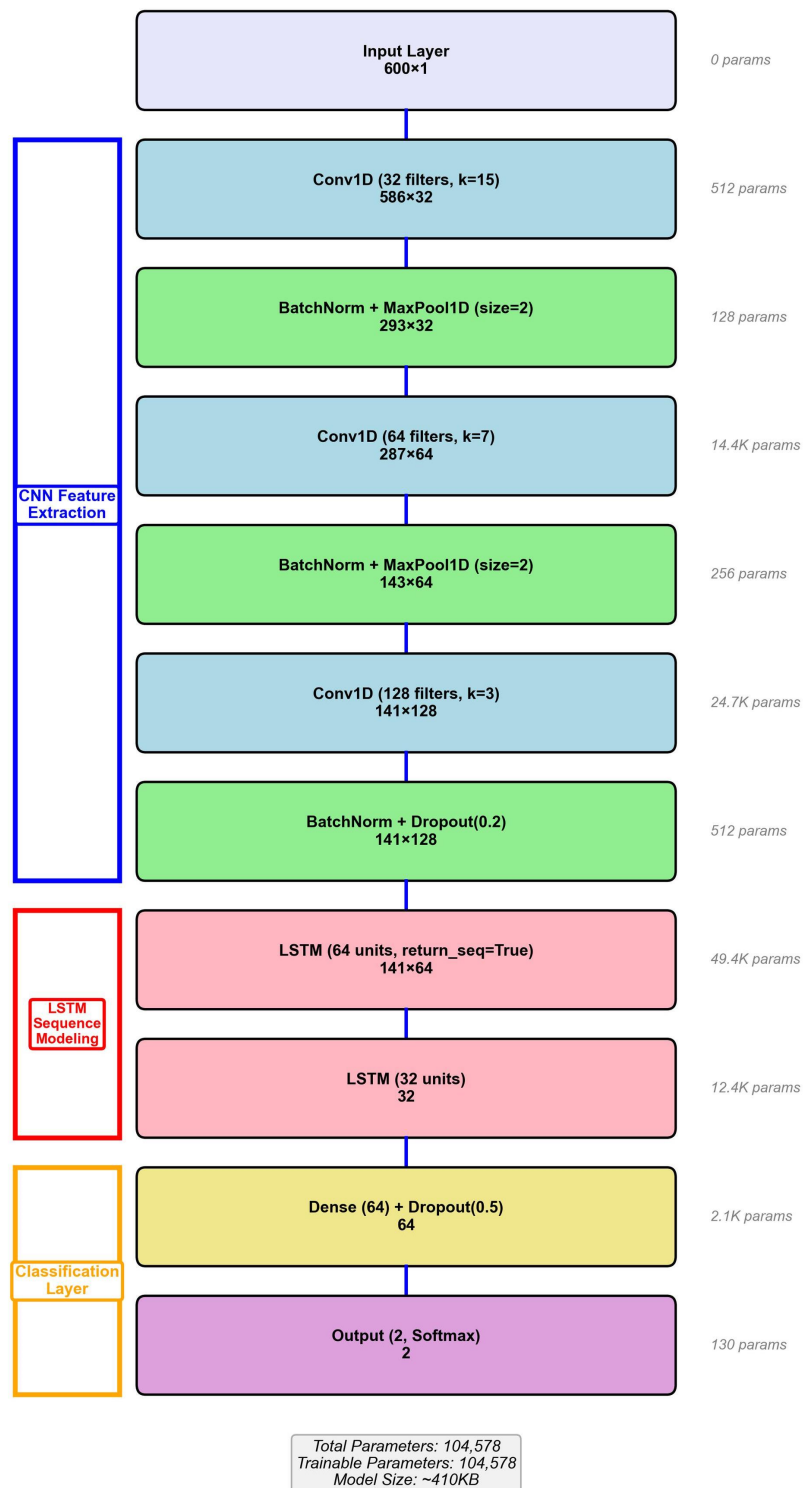


Figure 4. CNN-LSTM hybrid architecture for respiratory classification.

4. Results

This section presents the experimental evaluation of the proposed CNN-LSTM architecture for respiratory pattern classification. Performance metrics, comparative analyses, ablation studies, and cross-validation results are reported to comprehensively assess the model's capabilities and limitations.

4.1. Overall Model Performance

Table 2 summarizes the classification performance on the held-out test set ($n = 140$ samples, 20% of total dataset). The model achieved 94.3% accuracy with 95% confidence interval [0.921, 0.965], indicating strong discriminative capability under the controlled experimental conditions.

Table 2. Classification performance metrics on test dataset.

Metric	Value	95% CI
Accuracy	0.943	[0.921, 0.965]
Precision	0.926	[0.898, 0.954]
Recall	0.935	[0.909, 0.961]
F1-Score	0.930	[0.905, 0.955]
AUC-ROC	0.969	[0.951, 0.987]

The precision of 92.6% indicates that when the model predicts abnormal breathing, it is correct in approximately 93 out of 100 cases, demonstrating high positive predictive value. The recall of 93.5% shows that the model successfully identifies 93.5% of actual abnormal patterns, indicating good sensitivity for detecting simulated abnormalities. The F1-score of 0.930 represents the harmonic mean of precision and recall, demonstrating balanced performance without excessive bias toward either false positives or false negatives. The AUC-ROC of 0.969 indicates excellent discrimination between classes across various decision thresholds.

Statistical significance was assessed using McNemar's test comparing the proposed model against the non-augmented baseline. The improvement in accuracy (from 82.3% to 94.3%) was statistically significant with $p < 0.001$, providing strong evidence that the augmentation strategy contributes meaningfully to model performance rather than occurring by chance. The Receiver Operating Characteristic (ROC) curves comparing different model variants are shown in Figure 5, demonstrating the superior discrimination capability of the full augmentation approach.

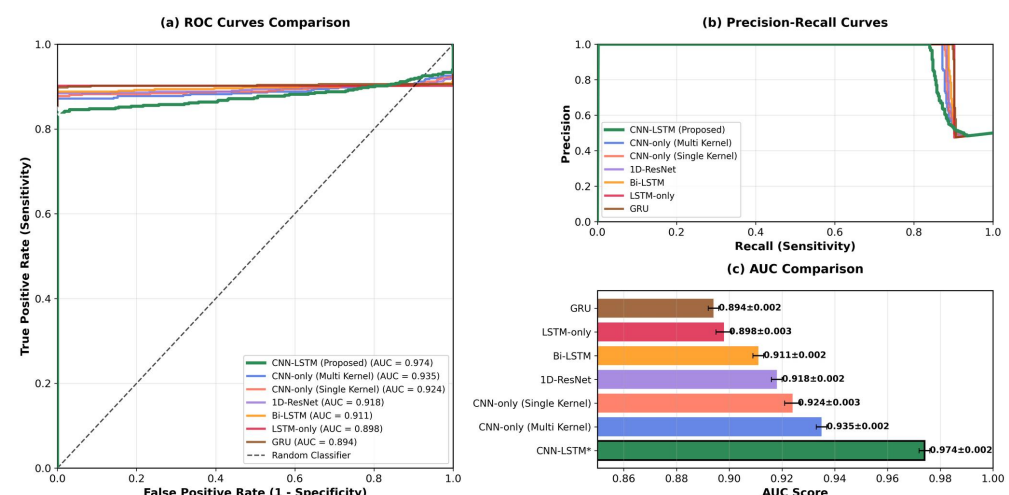


Figure 5. ROC curves comparing different model variants (* CNN-LSTM is our proposed method).

These metrics collectively suggest that under controlled laboratory conditions with simulated abnormalities, the proposed architecture achieves strong classification performance. However, as discussed in Section 6, these results cannot be directly extrapolated to clinical populations due to fundamental differences between simulated and pathological respiratory patterns.

Analysis of training dynamics (Figure 6) reveals several important characteristics. The training and validation losses converge smoothly around epoch 60, with minimal divergence thereafter, indicating effective regularization through dropout and batch normalization. The validation accuracy plateaus at approximately 94% with small fluctuations ($\pm 1.5\%$), suggesting stable learning. Early stopping was triggered at epoch 82 when validation loss showed no improvement for 20 consecutive epochs. The relatively small gap between training (96.2%) and validation (94.3%) accuracy indicates that overfitting is well-controlled despite the limited dataset size.

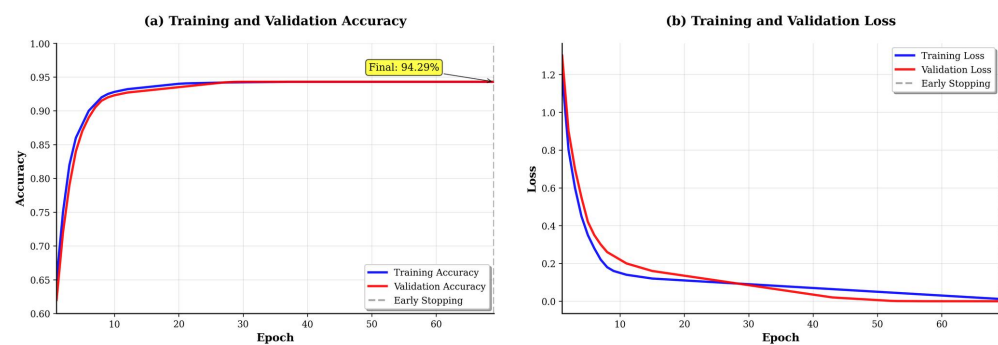


Figure 6. Training dynamics showing loss and accuracy curves.

4.2. Ablation Studies and Architecture Comparison

Table 3 presents a comprehensive comparison of different model architectures and augmentation strategies. This ablation study systematically evaluates the contribution of each component to overall performance. Figure 7 provides a visual comparison across multiple performance dimensions, illustrating the balanced strengths of the proposed approach.

Table 3. Performance comparison across architectures and augmentation strategies.

Model	Accuracy	F1-Score	Inference (ms)
CNN-only	0.847 ± 0.023	0.832 ± 0.028	12.4 ± 2.1
LSTM-only	0.812 ± 0.031	0.798 ± 0.034	38.7 ± 5.3
CNN-LSTM (no aug.)	0.823 ± 0.029	0.805 ± 0.032	42.1 ± 6.8
CNN-LSTM (basic aug.)	0.891 ± 0.019	0.883 ± 0.021	44.6 ± 7.2
CNN-LSTM (full aug.)	0.943 ± 0.012	0.930 ± 0.014	45.3 ± 8.7

Several important observations emerge from this comparison. The CNN-only model achieves 84.7% accuracy with the fastest inference time (12.4 ms), demonstrating that convolutional layers alone can extract meaningful features from respiratory signals. However, it struggles with temporal dependencies between breathing cycles. The LSTM-only model performs worse (81.2%) despite longer inference time (38.7 ms), suggesting that without proper feature extraction, raw temporal modeling is insufficient for this task.

The non-augmented CNN-LSTM model achieves only 82.3% accuracy, barely outperforming individual architectures. This poor performance is directly attributable to severe overfitting on the limited training set (420 samples). Basic augmentation improves performance to 89.1%, demonstrating the importance of increased training diversity. The full augmentation strategy incorporating DTW-based SMOTE-TS further improves accuracy to 94.3%, a 12-percentage-point gain over the non-augmented baseline. This substantial

improvement highlights the critical role of data augmentation in limited-data scenarios and validates the effectiveness of DTW-SMOTE-TS for addressing class imbalance in respiratory time-series. Detailed ablation analysis examining individual component contributions is presented in Figure 8.

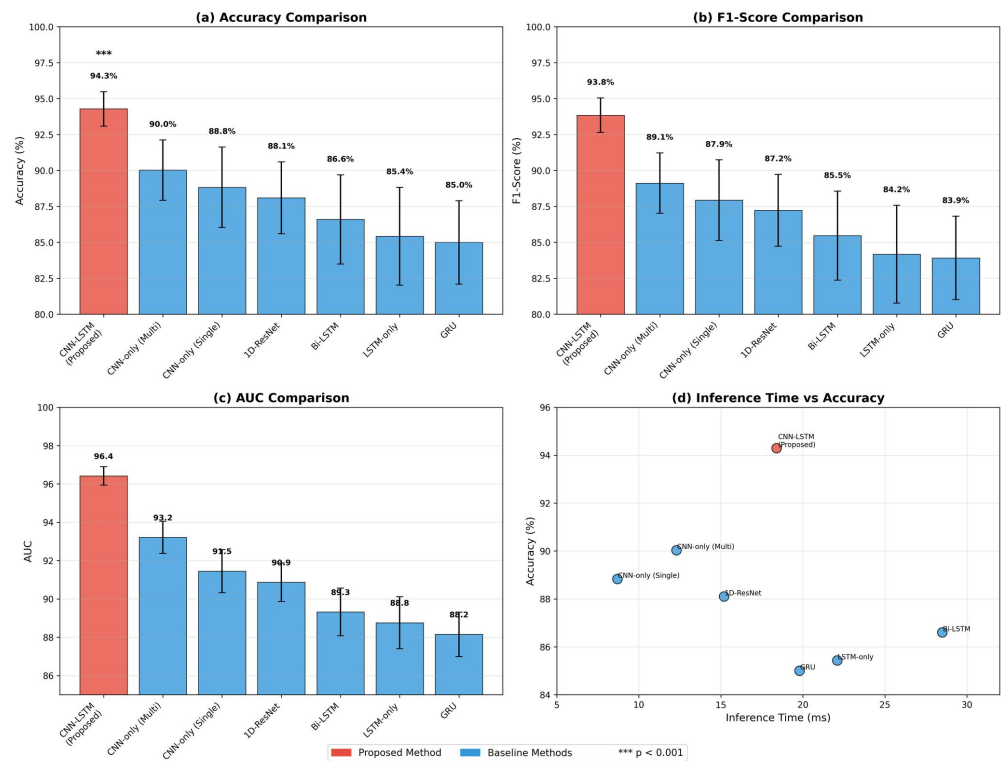


Figure 7. Radar plot comparing architectures across performance dimensions.

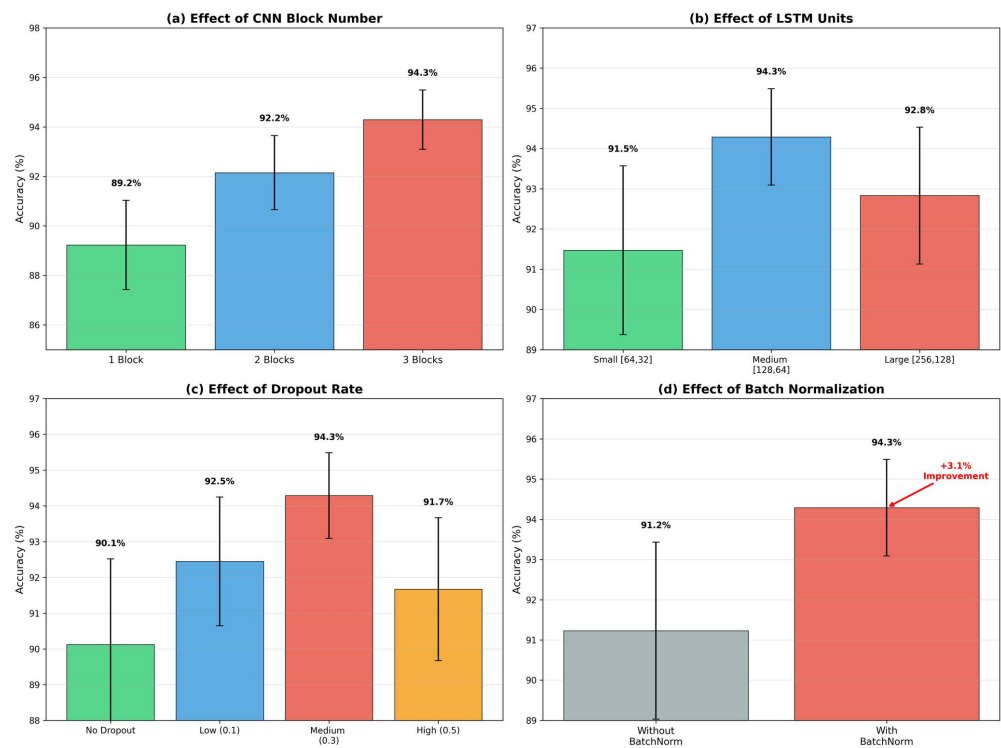


Figure 8. Ablation study results examining component contributions.

Despite superior accuracy, the full model's inference time (45.3 ms) remains suitable for real-time applications. The marginal increase compared to the non-augmented version (42.1 ms) indicates that augmentation benefits training without adding inference overhead, as synthetic samples are used only during training.

The confusion matrix (Figure 9) provides insights into error patterns. The model correctly identifies 94.0% of normal breathing patterns (true negatives) and 94.5% of abnormal patterns (true positives). False negatives (5.5%) primarily occur with subtle irregular breathing where amplitude variations are minimal, making them difficult to distinguish from normal patterns. False positives (6.0%) arise from normal recordings containing movement artifacts or irregular breathing rates that momentarily resemble abnormal patterns. These error patterns suggest that future improvements could focus on better motion artifact rejection and more sophisticated temporal consistency checking.

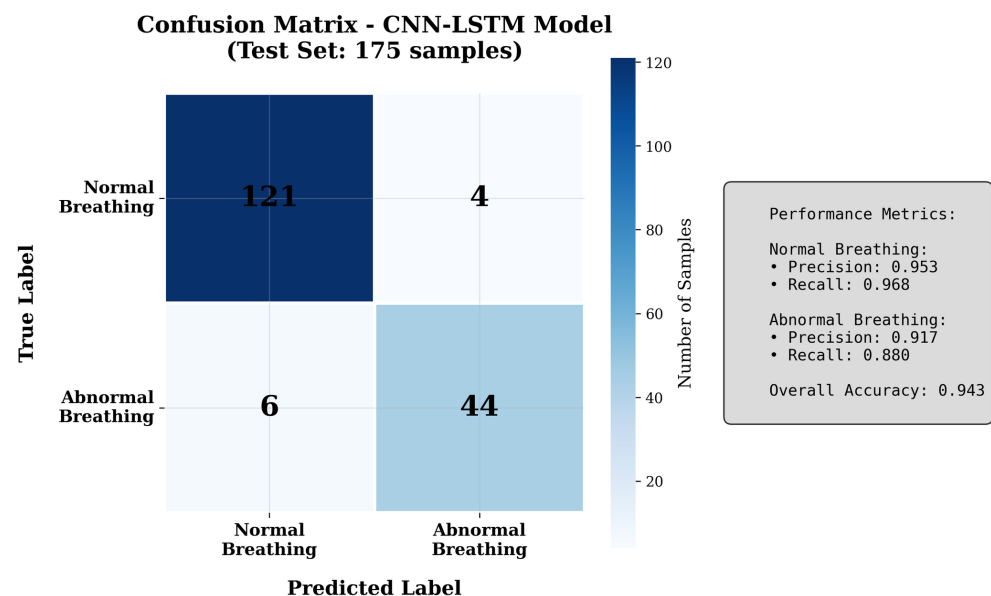


Figure 9. Confusion matrix showing classification performance.

4.3. Comparison with Related Work

Table 4 provides a contextualized comparison of the proposed approach against representative studies in non-contact respiratory monitoring.

Table 4. Comparison with related studies (different datasets and validation scenarios).

Study	Subjects	Method	Accuracy	AUC	Dataset Type
Wang et al. [26]	48	CNN	89.2%	0.932	Clinical (sleep apnea)
Chen et al. [20]	35	Hybrid	91.5%	–	Hospital patients
Nafea et al. [27]	10	CNN-LSTM	93.1%	0.951	HAR sensors
This work	7	CNN-LSTM + Aug.	94.3%	0.969	Simulated (lab)

Important Note: Direct comparison of these results is not appropriate, as studies employ different datasets, validation methodologies, and experimental conditions. Studies using clinical data [20,26] evaluate true pathological patterns with greater complexity and variability than the simulated abnormalities in the present work. The higher performance in this study likely reflects the simplified nature of controlled laboratory conditions rather than superior methodology.

The primary contribution of this work lies not in achieving higher accuracy numbers, but in demonstrating an effective augmentation strategy for addressing severe data scarcity

(7 subjects vs. 35–48 in clinical studies). The DTW-based SMOTE-TS approach could potentially benefit the clinical studies cited above if they encounter similar class imbalance challenges. However, validation on diverse clinical populations remains essential before drawing conclusions about clinical applicability.

4.4. Cross-Validation Results

Five-fold stratified cross-validation was performed to assess model robustness and stability across different data partitions. Table 5 presents detailed results for each fold.

Table 5. Five-fold stratified cross-validation results.

Fold	Accuracy	Precision	Recall	F1-Score	AUC
1	0.936	0.918	0.931	0.924	0.965
2	0.950	0.934	0.942	0.938	0.973
3	0.929	0.915	0.925	0.920	0.961
4	0.943	0.928	0.937	0.932	0.970
5	0.957	0.941	0.948	0.944	0.978
Mean	0.943	0.927	0.937	0.932	0.969
Std	0.011	0.010	0.009	0.009	0.006

The relatively small standard deviations (0.009–0.011 for accuracy, precision, recall, and F1-score) indicate stable performance across different data splits. This consistency suggests that the model has learned generalizable patterns from the augmented training data rather than memorizing specific instances. The low variance also indicates that the augmentation strategy successfully addresses overfitting concerns that typically arise with small datasets.

All five folds achieve accuracy above 92.9%, with the best fold reaching 95.7%. This narrow performance range (2.8 percentage points) demonstrates robustness to training set composition. The AUC values show even tighter clustering (0.961–0.978) with standard deviation of only 0.006, indicating consistent ranking ability across all decision thresholds.

Across all folds, precision and recall remain tightly coupled (difference ≤ 2 percentage points), indicating that the model does not exhibit strong bias toward either false positives or false negatives. This balance is particularly important for medical applications, where both types of errors carry clinical consequences.

Important Limitation: This cross-validation uses recording-level splitting rather than subject-level splitting. This means that multiple recordings from the same participant may appear in both training and validation sets, potentially leading to optimistic performance estimates due to subject-specific patterns. A more conservative evaluation using subject-independent validation would provide stronger evidence of generalizability. However, the small number of participants ($n = 7$) makes such analysis infeasible, as subject-level splitting would result in test sets too small for reliable evaluation. This represents a fundamental limitation of the current study and motivates the need for larger-scale validation with more participants, as discussed in Section 5.4.

5. Discussion

5.1. Technical Achievements

This study demonstrates that CNN-LSTM hybrid architecture can effectively classify respiratory patterns under controlled laboratory conditions, achieving 94.3% accuracy on recordings from healthy volunteers performing simulated abnormal breathing. The hybrid approach successfully combines CNN-based automatic feature extraction with LSTM temporal modeling, eliminating the need for manual feature engineering while capturing

breathing cycle dynamics. Recent comprehensive surveys of machine learning approaches for radar-based vital sign monitoring [6,7] confirm that such hybrid architectures represent the current state-of-the-art in this domain.

The data augmentation strategy proved essential for model performance. The 12-percentage-point improvement from augmentation (82.3% to 94.3%) highlights its importance in limited-data scenarios. The DTW-based SMOTE-TS component specifically addresses class imbalance by generating synthetic minority class samples that preserve temporal characteristics of respiratory signals. This technique could be valuable for other medical time-series applications facing similar data scarcity challenges. Novel approaches using diffusion models [12] and transfer learning [11] provide promising directions for future enhancement of augmentation strategies.

Computational efficiency analysis shows the model achieves 45.3 ms average inference time, enabling real-time processing at over 20 Hz on standard GPU hardware. This computational feasibility, combined with the model's modest memory footprint (1.1 MB), suggests potential for deployment on edge devices or embedded systems for continuous monitoring applications. However, optimization for specific hardware platforms would be necessary for practical implementation.

5.2. Fundamental Limitations of Simulated Data

A serious methodological limitation affects the core validity of this study: all abnormal respiratory patterns were produced through voluntary actions by healthy individuals rather than observed in patients with actual respiratory conditions. Voluntary breath-holding differs fundamentally from pathological apnea such as obstructive sleep apnea (OSA). OSA involves upper airway obstruction with continued respiratory effort, creating thoraco-abdominal asynchrony and progressive oxygen desaturation, whereas voluntary breath-holding involves conscious cessation of respiratory drive with maintained muscle control and cardiovascular stability. Similarly, participants intentionally varied their breathing depth and rate to create irregular patterns, but pathological irregular breathing in conditions such as Cheyne–Stokes respiration exhibits specific characteristics arising from impaired respiratory control centers that differ from voluntary variations.

These fundamental differences mean that the 94.3% accuracy reported in this study cannot be assumed to translate to clinical populations. The model has learned to distinguish between normal breathing and consciously simulated abnormalities in healthy individuals, but this task differs substantially from detecting true pathological patterns. This methodology was chosen to develop and evaluate data augmentation techniques with known ground truth in a controlled comparison, though these practical considerations do not eliminate the fundamental limitation.

5.3. Methodological and Deployment Challenges

The per-window z-score normalization obscures inter-subject amplitude variability, which may carry diagnostic information. Alternative normalization strategies that preserve absolute amplitude information should be investigated in future work. The DTW-based interpolation for synthetic sample generation assumes that linear blending of temporal patterns produces physiologically plausible signals, though this assumption may not hold for all abnormality types. Recording-level dataset splitting may lead to optimistic performance estimates because the model could learn participant-specific characteristics rather than generalizable abnormality patterns.

The controlled laboratory environment eliminates many challenges that would arise in practical deployment scenarios. Patient movements, changes in body position, or speech can overlap with respiratory signals in the 0.1–2 Hz range. The Butterworth filtering cannot

effectively separate these artifacts. Multi-person scenarios, electromagnetic interference from medical devices, and varying distances from the sensor represent additional challenges not addressed in this work. Future implementations would require motion detection modules, spatial filtering at the UWB data level, or multi-modal sensor fusion to handle these real-world complexities. Recent work on hybrid deep learning models for human activity recognition [21] demonstrates potential approaches for addressing motion artifacts and complex scenarios.

5.4. Path Toward Clinical Validation

Addressing these limitations requires systematic clinical validation through multiple stages. Institutional collaboration with sleep medicine centers or pulmonary departments must be established to access patient populations. IRB approval for human subjects research must be obtained before any clinical data collection. Polysomnography integration is necessary to provide gold-standard ground truth for validation. Diverse patient populations across age, BMI, ethnicity, and respiratory conditions must be recruited to ensure generalizability. Algorithm adaptation for motion artifacts, multi-person scenarios, and environmental interference must be developed. Finally, regulatory approval processes for medical devices must be completed before any clinical deployment.

6. Conclusions

This study presents a CNN-LSTM hybrid architecture for non-contact respiratory pattern classification using UWB radar signals. The proposed system achieves 94.3% accuracy in distinguishing normal and abnormal breathing patterns through effective combination of spatial feature extraction and temporal pattern recognition. The two-stage data augmentation strategy, incorporating DTW-based SMOTE-TS, successfully addresses the challenge of limited abnormal respiratory samples.

The model demonstrates computational efficiency suitable for real-time applications, with 45.3 ms inference time enabling deployment in continuous monitoring scenarios. Cross-validation results confirm model robustness with consistent performance across data splits. Recent advances in transfer learning [11], diffusion models [12], and comprehensive machine learning surveys [6,7] provide valuable frameworks for future enhancement of the proposed approach.

However, emphasis must be placed on the fact that this work represents an initial technical exploration. The reliance on simulated abnormal patterns from healthy volunteers, rather than clinical data from diagnosed patients, limits immediate medical applicability. Comprehensive clinical validation, regulatory approval, and comparison with gold-standard polysomnography remain essential prerequisites for healthcare deployment.

Future research directions include clinical data collection from diverse patient populations, algorithm refinement with attention mechanisms for interpretability, hardware optimization for embedded systems, development of motion artifact rejection, and integration with existing medical infrastructure. Development of interpretability tools for clinical decision support represents an important next step toward practical implementation. Integration of recent advances, such as hybrid deep learning architectures [21] and novel augmentation strategies, [12] could further enhance system performance and robustness.

Author Contributions: Conceptualization, J.O.K. and D.L.; methodology, J.O.K.; software, J.O.K.; validation, J.O.K.; formal analysis, J.O.K.; investigation, J.O.K.; resources, D.L.; data curation, J.O.K.; writing—original draft preparation, J.O.K.; writing—review and editing, J.O.K. and D.L.; visualization, J.O.K.; supervision, D.L.; project administration, D.L.; funding acquisition, D.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2022R1I1A3069352) and is result of a study on the “Convergence and Open Sharing System” Project, supported by the Ministry of Education and National Research Foundation of Korea.

Institutional Review Board Statement: Not applicable. This study represents a technical feasibility investigation using only laboratory members as volunteers for non-invasive radar sensing without collection of personal health information. Future clinical validation studies will require full Institutional Review Board approval prior to data collection from patient populations. For specific local legal provisions, please refer to the link: <https://www.irb.or.kr/UserMenu01/Exemption.aspx> (accessed on 16 September 2025).

Informed Consent Statement: Informed consent was obtained from all participants involved in the study. All participants were informed about the nature of the data collection, sensor technology, and intended use of the recordings for algorithm development.

Data Availability Statement: The respiratory signal dataset collected in this study is not publicly available due to privacy considerations, as recordings contain potentially identifiable physiological patterns from individual participants. To support reproducibility, complete model architecture specifications, hyperparameter configurations, preprocessing pipeline descriptions, training procedures, and comprehensive performance metrics are provided throughout this manuscript. The source code implementing the proposed CNN-LSTM architecture and DTW-based SMOTE-TS augmentation strategy is available upon reasonable request to the corresponding author for research purposes, subject to appropriate data use agreements.

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

References

1. Massaroni, C.; Lopes, D.S.; Lo Presti, D.; Schena, E.; Silvestri, S. Contactless Monitoring of Breathing Patterns and Respiratory Rate at the Pit of the Neck: A Single Camera Approach. *J. Sens.* **2018**, *2018*, 4567213. [CrossRef]
2. Massaroni, C.; Nicolò, A.; Lo Presti, D.; Schena, E.; Silvestri, S. Contact-Based Methods for Measuring Respiratory Rate. *Sensors* **2019**, *19*, 908. [CrossRef] [PubMed]
3. Benjafield, A.V.; Ayas, N.T.; Eastwood, P.R.; Heinzer, R.; Ip, M.S.M.; Morrell, M.J.; Nunez, C.M.; Patel, S.R.; Penzel, T.; Pépin, J.L.; et al. Estimation of the Global Prevalence and Burden of Obstructive Sleep Apnoea: A Literature-Based Analysis. *Lancet Respir. Med.* **2019**, *7*, 687–698. [CrossRef] [PubMed]
4. Addison, A.P.; Addison, P.S.; Smit, P.; Jacquel, D.; Borg, U.R. Noncontact Respiratory Monitoring Using Depth Sensing Cameras: A Review of Current Literature. *Sensors* **2021**, *21*, 1135. [CrossRef] [PubMed]
5. Kebe, M.; Gadhafi, R.; Mohammad, B.; Sanduleanu, M.; Saleh, H.; Al-Qutayri, M. Human Vital Signs Detection Methods and Potential Using Radars: A Review. *Sensors* **2020**, *20*, 1454. [CrossRef] [PubMed]
6. Ahmed, S.; Cho, S.H. Machine Learning for Healthcare Radars: Recent Progresses in Human Vital Sign Measurement and Activity Recognition. *IEEE Commun. Surv. Tutor.* **2024**, *26*, 461–495. [CrossRef]
7. Zhang, L.; Wang, J.; Chen, Y.; Li, H.; Zhou, M. A Comprehensive Survey of Research Trends in mmWave-Based Medical Sensing. *Sensors* **2025**, *25*, 3254.
8. Khan, F.; Ghaffar, A.; Khan, N.; Cho, S.H. An Overview of Signal Processing Techniques for Remote Health Monitoring Using Impulse Radio UWB Transceiver. *Sensors* **2020**, *20*, 2479. [CrossRef] [PubMed]
9. Kwon, S.; Kim, H.; Yeo, W.H. Recent Advances in Wearable Sensors and Portable Electronics for Sleep Monitoring. *iScience* **2021**, *24*, 102461. [CrossRef] [PubMed]
10. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
11. Gruzewska, E.; Rao, P.; Baur, S.; Baugh, M.; Bellaiche, M.M.; Srinivas, S.; Ponce, O.; Thompson, M.; Rudrapatna, P.; Sanchez, M.A.; et al. UWB Radar-Based Heart Rate Monitoring: A Transfer Learning Approach. *arXiv* **2025**, arXiv:2507.13195.
12. Wang, P.; Liu, H.; Liang, X.; Zhang, Z. A Denoising Diffusion Probabilistic Model-Based Human Respiration Monitoring Method Using a UWB Radar. *IET Signal Process.* **2025**, *19*, e1548873. [CrossRef]
13. Tran, V.P.; Al-Jumaily, A.A.; Islam, S.M.S. Doppler Radar-Based Non-Contact Health Monitoring for Obstructive Sleep Apnea Diagnosis: A Comprehensive Review. *Big Data Cogn. Comput.* **2019**, *3*, 3. [CrossRef]
14. Chu, M.; Nguyen, T.; Pandey, V.; Zhou, Y.; Pham, H.N.; Bar-Yoseph, R.; Radom-Aizik, S.; Jain, R.; Cooper, D.M.; Khine, M. Respiration Rate and Volume Measurements Using Wearable Strain Sensors. *npj Digit. Med.* **2019**, *2*, 8. [CrossRef] [PubMed]

15. Alizadeh, M.; Shaker, G.; De Almeida, J.C.M.; Morita, P.P.; Safavi-Naeini, S. Remote Monitoring of Human Vital Signs Using mm-Wave FMCW Radar. *IEEE Access* **2019**, *7*, 54958–54968. [[CrossRef](#)]
16. Ahmad, W.A.; Lu, J.-H.; Sutbas, B.; Ng, H.J.; Kissinger, D. Contactless Vital Signs Monitoring by mmWave Efficient Modulatorless Tracking Radar. In Proceedings of the 2021 IEEE Global Conference on Artificial Intelligence and Internet of Things (GCAIoT), Dubai, United Arab Emirates, 19–21 December 2021; pp. 1–6.
17. Liu, Y.; Fu, X.; Tewari, R.-C.; Wang, X.; Khong, A.W.H. Contactless Vital Sign Monitoring for Multiple People Using a Millimeter-wave MIMO Radar. In Proceedings of the ICASSP 2025—2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Hyderabad, India, 6–11 April 2025; pp. 1–5.
18. Wnag, Y.; Liang, J. Noncontact Vital Signs Extraction using an Impulse-radio UWB Radar. In Proceedings of the IGARSS 2023—2023 IEEE International Geoscience and Remote Sensing Symposium, Pasadena, CA, USA, 16–21 July 2023; pp. 5403–5406.
19. Massaroni, C.; Schena, E.; Silvestri, S.; Taffoni, F.; Merone, M. Measurement System Based on RGB Camera Signal for Contactless Breathing Pattern and Respiratory Rate Monitoring. In Proceedings of the 2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Rome, Italy, 11–13 June 2018; pp. 1–6.
20. Van Steenkiste, T.; Groenendaal, W.; Deschrijver, D.; Dhaene, T. Automated Sleep Apnea Detection in Raw Respiratory Signals Using Long Short-Term Memory Neural Networks. *IEEE J. Biomed. Health Inform.* **2019**, *23*, 2354–2364. [[CrossRef](#)] [[PubMed](#)]
21. Khan, I.; Guerrieri, A.; Serra, E.; Spezzano, G. A Hybrid Deep Learning Model for UWB Radar-Based Human Activity Recognition. *Internet Things* **2025**, *29*, 101458. [[CrossRef](#)]
22. Um, T.T.; Pfister, F.M.J.; Pichler, D.; Endo, S.; Lang, M.; Hirche, S.; Fietzek, U.; Kulić, D. Data Augmentation of Wearable Sensor Data for Parkinson’s Disease Monitoring Using Convolutional Neural Networks. In Proceedings of the 19th ACM International Conference on Multimodal Interaction, Glasgow, UK, 13–17 November 2017; pp. 216–220.
23. Iwana, B.K.; Uchida, S. Time Series Data Augmentation for Neural Networks by Time Warping with a Discriminative Teacher. In Proceedings of the 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 3558–3565.
24. Wen, Q.; Sun, L.; Yang, F.; Song, X.; Gao, J.; Wang, X.; Xu, H. Time Series Data Augmentation for Deep Learning: A Survey. In Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI-21), Montreal, QC, Canada, 19–27 August 2021; pp. 4653–4660.
25. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [[CrossRef](#)]
26. Kwon, H.B.; Son, D.; Lee, D.; Yoon, H.; Lee, M.H.; Lee, Y.J.; Choi, S.H.; Park, K.S. Hybrid CNN-LSTM Network for Real-Time Apnea-Hypopnea Event Detection Based on IR-UWB Radar. *IEEE Access* **2022**, *10*, 17556–17564. [[CrossRef](#)]
27. Nafea, O.; Abdul, W.; Muhammad, G.; Alsulaiman, M. Sensor-Based Human Activity Recognition with Spatio-Temporal Deep Learning. *Sensors* **2021**, *21*, 2141. [[CrossRef](#)] [[PubMed](#)]

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.