

Lightweight Federated Deep Learning for COPD Exacerbation Risk Prediction Using Multimodal Ubiquitous Health Sensing

Riad Alharbey

Department of Information Systems and Technology, College of Computer Science and Engineering,
University of Jeddah, Jeddah 21959, Saudi Arabia

Abstract—Chronic obstructive pulmonary disease (COPD) exacerbations are difficult to anticipate because deterioration develops across heterogeneous respiratory, physiological, behavioral, environmental, and clinical signals. This study presents FedSqueeze-HGRUViTNet, a lightweight federated deep learning model for COPD exacerbation-risk prediction using multimodal wearable and harmonized clinical information. The model integrates depthwise separable convolution for local health-index variation, a health-gated residual gated recurrent unit for patient-specific temporal deterioration, and a squeeze-enhanced Health Vision Transformer for broader multimodal interactions. Personalized baseline deviation, explicit missingness indicators, squeeze-based recalibration, adaptive feature fusion, and weighted federated averaging are incorporated to support distributed learning without centralizing raw patient records. The model was evaluated on the TOLIFE COPD Health Indices Dataset and a harmonized Medical Information Mart for Intensive Care IV (MIMIC-IV) COPD cohort using patient-level 70:10:20 training, validation, and testing partitions. Each experiment was repeated across five random seeds using fixed patient partitions. On TOLIFE, FedSqueeze-HGRUViTNet achieved 97.30% accuracy and an area under the receiver operating characteristic curve (AUC) of 0.985 for binary exacerbation-risk prediction, together with 96.72% accuracy for three-class risk stratification. On the complementary MIMIC-IV electronic health record-based evaluation, the model achieved 96.62% accuracy and an AUC of 0.978. Robustness experiments under missing observations, Gaussian-noise perturbation, monitoring-window variation, and non-independent and non-identically distributed client conditions showed limited performance degradation, with accuracy remaining above 95.6% across the evaluated settings. Inference required 5.18 ms per sample in the evaluated computing environment. These results indicate that FedSqueeze-HGRUViTNet provides an accurate and computationally efficient federated approach for distributed COPD exacerbation-risk assessment across heterogeneous multimodal health information.

Keywords—COPD exacerbation prediction; federated learning; wearable health monitoring; multimodal learning; ubiquitous sensing; deep learning

I. INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a progressive respiratory disorder associated with persistent airflow limitation, functional decline, and recurrent exacerbations. These events accelerate disease progression, increase hospitalization and readmission, and reduce quality of life. Early prediction remains challenging because deterioration reflects interacting respiratory, physiological, behavioral, environmen-

tal, and clinical factors [1]. Conventional assessment through clinical examination, spirometry, symptoms, and laboratory testing is episodic and provides limited coverage of gradual changes preceding an exacerbation. Longitudinal multimodal monitoring provides complementary information from activity, sleep, respiratory behavior, physiological stability, mobility, environmental exposure, and deviations from a patient's usual health state [2].

Artificial intelligence (AI) and machine-learning methods have increasingly been applied to COPD exacerbation prediction and digital respiratory monitoring [3]. A single physiological or clinical source does not fully represent the interacting patterns of COPD deterioration [4]. Wearable and clinical data are also frequently incomplete, noisy, irregularly sampled, imbalanced, and influenced by patient-specific baselines. Interpretable machine-learning approaches have therefore been used to identify influential COPD-related health characteristics [5]. Smartphones, smartwatches, home spirometry, sleep-monitoring devices, mobility sensors, and environmental sensing technologies further support continuous COPD monitoring outside conventional clinical settings [6], [7]. Deep learning captures complementary patterns through convolutional, recurrent, and attention-based representations [8], although large hybrid architectures increase computational and inference costs and constrain resource-limited deployment.

COPD monitoring data are distributed across wearable platforms, clinics, and hospitals, while privacy and governance requirements restrict centralized sharing of raw patient records. Federated learning enables collaborative model training by exchanging model updates while retaining patient data at local computing nodes and has demonstrated potential in distributed healthcare applications [9], [10]. However, federated COPD exacerbation prediction from heterogeneous multimodal health indices remains insufficiently explored under missing measurements, patient variability, and non-independent and non-identically distributed (non-IID) client data.

To address these challenges, this study proposes FedSqueeze-HGRUViTNet, a lightweight federated deep learning model for COPD exacerbation-risk prediction. The model integrates depthwise separable convolution for local health-index variation, a health-gated residual gated recurrent unit (HGRU) for patient-specific deterioration progression, and a squeeze-enhanced Health Vision Transformer for broader multimodal interactions. Personalized baseline

deviation, missingness indicators, squeeze-based recalibration, and adaptive fusion are incorporated, while weighted federated averaging supports distributed optimization without centralizing raw patient records. The model is evaluated on the TOLIFE COPD Health Indices Dataset and a harmonized Medical Information Mart for Intensive Care IV (MIMIC-IV) COPD cohort using binary and three-class prediction, computational-efficiency analysis, robustness testing, client-heterogeneity analysis, and centralized-versus-federated comparisons. The main contributions are summarized as follows:

- A lightweight federated model, FedSqueeze-HGRUViTNet, is developed for COPD exacerbation-risk prediction from heterogeneous respiratory, physiological, behavioral, environmental, and clinical health indices while retaining patient data at local computing nodes.
- A complementary three-branch architecture combines depthwise separable convolution, health-gated residual recurrent learning, and squeeze-enhanced Health Vision Transformer attention to capture local variation, patient-specific deterioration, and global multimodal interactions.
- A personalized representation and adaptive fusion strategy integrates patient-specific baseline deviation, missingness information, squeeze recalibration, and dynamic branch weighting to improve learning from incomplete and heterogeneous monitoring data.
- The model is evaluated on TOLIFE and harmonized MIMIC-IV data through binary and three-class prediction, computational-efficiency and robustness assessment, and federated experiments under heterogeneous client distributions.

The remainder of the paper presents related work, the proposed methodology and datasets, experimental results, discussion, and conclusions.

II. RELATED WORK

Recent COPD research has applied deep learning, multimodal representation learning, wearable sensing, medical imaging, and digital-health monitoring to disease assessment and exacerbation prediction. Existing work mainly follows two complementary directions: representation learning from physiological, respiratory, imaging, and clinical data, and prediction of exacerbation, deterioration, or related clinical outcomes.

A. Deep Feature Representation for COPD Monitoring

Deep learning has been used to extract COPD-related patterns from heterogeneous data sources. DeepSpiro learned respiratory-function representations from spirogram time series for COPD detection and future-risk prediction [11], while AI-PFT-Clin combined pulmonary-function curve images with clinical information for exacerbation prediction [12]. COPD-MMDDxNet integrated electronic medical record data through multimodal representation and dynamic fusion for COPD diagnosis [13]. These studies demonstrate the value of respiratory and clinical information, although their inputs are primarily

structured pulmonary-function or hospital-based data rather than continuous multimodal wearable monitoring.

Transformer and hybrid architectures have also been investigated. COPD-TransNet combined a Swin Transformer with quantitative emphysema features from computed tomography (CT) images for COPD detection and staging [14]. A hybrid convolutional neural network (CNN), bidirectional long short-term memory (BiLSTM), and Vision Transformer (ViT) architecture was applied to respiratory-sound-based COPD detection [8], while a CNN-recurrent neural network (RNN) model learned spatial and sequential representations from lung sounds [16]. Smartphone-based cough-sound analysis has also been investigated for accessible COPD detection [15]. These approaches demonstrate the usefulness of convolutional, recurrent, and attention-based representations, but their primary focus is disease detection or classification rather than longitudinal exacerbation-risk prediction from multimodal health-index streams.

B. Exacerbation and Deterioration Prediction

Several studies have focused more directly on COPD exacerbation and deterioration assessment. An extreme gradient boosting model with Shapley additive explanations (XGBoost-SHAP) used clinical and laboratory variables for acute exacerbation-risk prediction [17]. A multimodal multitask framework combined CT imaging and clinical information for COPD identification and classification [18]. Respiratory-sound classification has also been investigated using a denoising-autoencoder and enhanced bidirectional LSTM framework with optimization (DAE-EBiLSTM-ASBO) [19], while PulmoClass-3DAtt combined three-dimensional imaging and self-attention for classification of COPD, preserved ratio impaired spirometry (PRISm), and normal cases [20]. Neural models using urine biomarkers have further demonstrated the predictive value of biological indicators for exacerbation-risk assessment [21].

Longitudinal and digital-health approaches are particularly relevant to early deterioration monitoring. A cyber-physical COPD management framework combined real-time health information with machine-learning models for risk estimation [22], while telemonitoring methods used symptom and monitoring histories to predict clinical deterioration [23]. Deep neural networks have also been applied to COPD readmission prediction after hospitalization [24], and inflammatory biomarkers have been used for acute exacerbation-risk modeling [25]. Interpretable machine learning based on breath volatile organic compounds has additionally been investigated for distinguishing COPD from related respiratory diseases. These approaches provide useful evidence for COPD prediction, but many rely on centralized processing, static clinical variables, laboratory measurements, or individual sensing modalities.

Overall, existing work demonstrates the value of spirometry, imaging, respiratory sounds, biomarkers, electronic health records, and telemonitoring for COPD assessment. However, limited work jointly addresses patient-specific baseline deviation, missingness-aware temporal progression, global multimodal interactions, lightweight computation, and federated learning within a single exacerbation-risk model. FedSqueeze-HGRUViTNet addresses this gap by integrating depthwise

separable convolution, health-gated residual recurrent learning, squeeze-enhanced Health Vision Transformer attention, adaptive feature fusion, and federated optimization.

III. MATERIALS AND METHODS

FedSqueeze-HGRUViTNet is a lightweight federated model for COPD exacerbation-risk prediction from respiratory, physiological, activity, sleep, mobility, environmental, demographic, and clinical information. The model combines personalized health-index representation, three complementary feature-learning branches, adaptive feature fusion, and federated optimization. Fig. 1 summarizes the complete workflow.

A. Datasets and Cohort Construction

The TOLIFE COPD Health Indices Dataset was used as the primary dataset for developing and evaluating FedSqueeze-HGRUViTNet [26]. It contains longitudinal multimodal COPD information collected from smartphones, smartwatches, spirometry, sleep and mobility sensors, environmental sensing, clinical examinations, and recorded exacerbation events. Clinical Study A included 74 participants, comprising 34 participants from Germany and 40 from Spain, monitored from May 2024 to May 2025.

After quality screening, temporal alignment, monitoring-window construction, and label assignment, 1,080,000 monitoring windows were retained. Participants were partitioned at the patient level into 52 training, 7 validation, and 15 testing participants, corresponding to 756,000 training, 108,000 validation, and 216,000 held-out testing windows. The resulting window-level distribution follows the 70:10:20 training, validation, and testing protocol. The prediction tasks included binary classification of stable versus upcoming exacerbation and three-class stratification into low-, moderate-, and high-risk states.

A COPD cohort derived from the Medical Information Mart for Intensive Care IV (MIMIC-IV) database was used for complementary EHR-based evaluation [27]. After cohort selection, temporal aggregation, feature harmonization, and sequence construction, 300,000 processed samples were retained and partitioned into 210,000 training, 30,000 validation, and 60,000 held-out testing samples. Because MIMIC-IV does not contain the same wearable sensing streams as TOLIFE, clinically compatible variables were organized into respiratory, physiological, demographic, laboratory/comorbidity, and clinical feature groups.

B. Preprocessing and Monitoring-Window Construction

The heterogeneous TOLIFE measurements were organized chronologically and synchronized to a common temporal representation. Sub-daily respiratory, physiological, mobility, sleep, and environmental variables were aligned at 10-minute intervals, while daily clinical and spirometry measurements were associated with the corresponding monitoring date. Records containing invalid timestamps, duplicated observations, or inadequate data-quality indicators were excluded.

Short missing intervals were handled using recent-value carry-forward, while longer gaps were imputed using medians estimated from the training partition. Extreme numerical values

were controlled through percentile-based winsorization, numerical variables were standardized using statistics estimated only from the training partition, and categorical variables were one-hot encoded. MIMIC-IV variables followed the same processing principles after hourly temporal aggregation.

Fixed-length TOLIFE sequences were constructed using a 14-day observation window. At 10-minute resolution, each monitoring window contained 2,016 time steps, and consecutive windows were generated using a 30-minute stride. Exacerbation records were used to assign stable, upcoming-exacerbation, and multi-level risk labels according to the temporal distance between each monitoring window and the next recorded exacerbation event.

Let $\Delta_{s,t}$ denote the number of days between monitoring window t and the next recorded exacerbation event for patient s . The three-class risk label is defined as

$$y_{s,t} = \begin{cases} 2, & 0 < \Delta_{s,t} \leq \tau_h, \\ 1, & \tau_h < \Delta_{s,t} \leq \tau_m, \\ 0, & \Delta_{s,t} > \tau_m, \end{cases} \quad (1)$$

where, classes 0, 1, and 2 denote low-, moderate-, and high-risk states, respectively. The thresholds were set to $\tau_h = 3$ days and $\tau_m = 7$ days, measured from the end of each observation window to the next recorded exacerbation. Binary prediction identified exacerbations occurring within the subsequent 7 days. MIMIC-IV labels represented COPD-related admission status.

C. Personalized Health-Index Representation

COPD patients exhibit different physiological and behavioral baselines; therefore, absolute measurements alone do not fully represent patient-specific deterioration. Patient-specific baseline profiles were estimated from observations designated as stable in the retrospective records. The normalized deviation of feature d from this baseline was calculated as

$$\delta_{s,t}^d = \frac{\tilde{x}_{s,t}^d - \mu_s^d}{\sigma_s^d + \epsilon}, \quad (2)$$

where, $\tilde{x}_{s,t}^d$ denotes the preprocessed value of feature d for patient s at time t , μ_s^d and σ_s^d denote the patient-specific baseline mean and standard deviation, and ϵ is a small numerical stability constant that prevents division by zero.

The final health-index representation combines the standardized observation, patient-specific deviation, and explicit missingness information:

$$\mathbf{e}_{s,t} = [\bar{\mathbf{x}}_{s,t}; \boldsymbol{\delta}_{s,t}; \mathbf{m}_{s,t}], \quad (3)$$

where, $\bar{\mathbf{x}}_{s,t}$ is the standardized feature vector, $\boldsymbol{\delta}_{s,t}$ contains patient-specific baseline deviations, and $\mathbf{m}_{s,t}$ is the corresponding missingness-indicator vector. This representation combines population-level health characteristics with individualized changes associated with COPD deterioration.

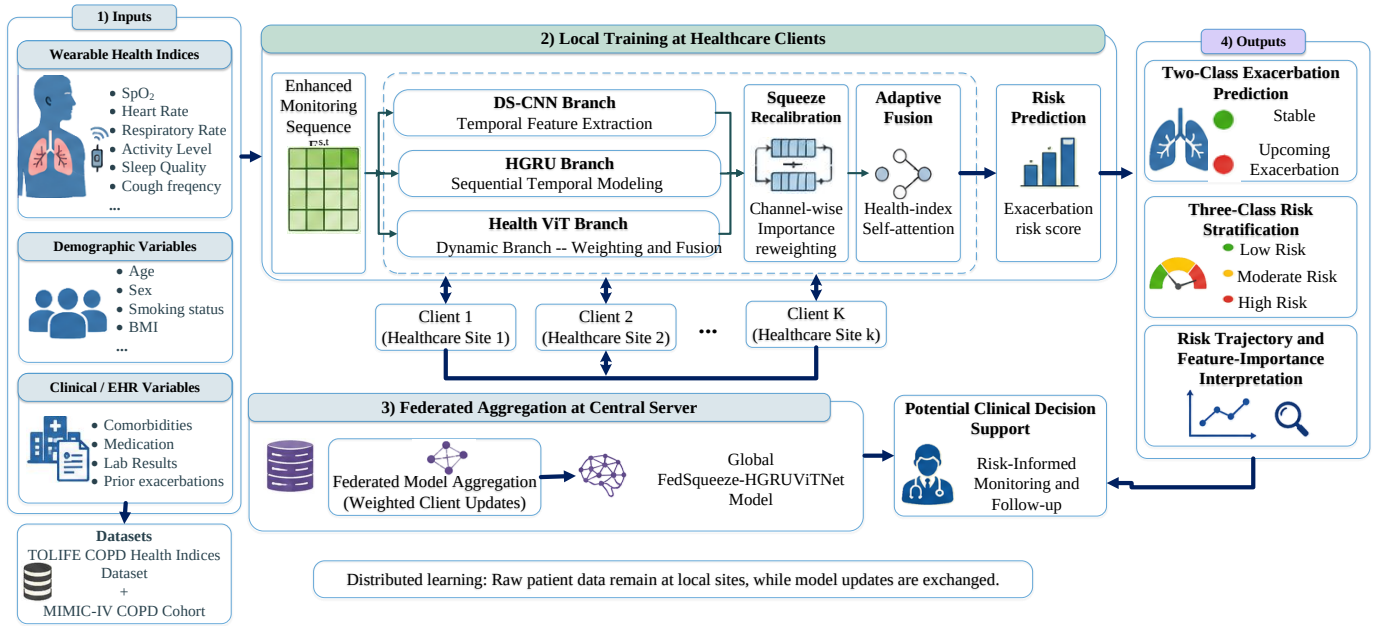


Fig. 1. Proposed FedSqueeze-HGRUViTNet model for COPD exacerbation-risk prediction using multimodal wearable, demographic, and clinical and electronic health record (EHR) inputs.

D. Proposed FedSqueeze-HGRUViTNet Architecture

FedSqueeze-HGRUViTNet learns complementary deterioration patterns through three parallel lightweight branches. The depthwise separable convolution branch captures short-range variations across respiratory, physiological, activity, sleep, mobility, and environmental health indices with reduced computational complexity. The health-gated residual gated recurrent unit (HGRU) branch models patient-specific temporal deterioration, while the squeeze-enhanced Health Vision Transformer captures broader interactions among heterogeneous health-information groups. The resulting representations are recalibrated and adaptively fused for exacerbation-risk prediction.

In the HGRU branch, the personalized health-index representation is processed through a health-deterioration gate:

$$\kappa_t = \sigma(\mathbf{A}_\kappa \mathbf{e}_{s,t} + \mathbf{b}_\kappa), \quad (4)$$

where, $\sigma(\cdot)$ denotes the sigmoid activation function, $\mathbf{A}_\kappa$ and $\mathbf{b}_\kappa$ are learnable gate parameters, and $\mathbf{e}_{s,t}$ contains the standardized observation, patient-specific deviation, and missingness information defined in Eq. 3. Together with the residual recurrent transition, the gate emphasizes temporal observations associated with stronger deterioration and produces the temporal representation $\mathbf{f}_{s,t}^h$.

The Health Vision Transformer organizes the monitoring sequence into group-aware representations covering respiratory, physiological, activity, sleep, mobility, environmental, and clinical information. Positional and health-group embeddings guide self-attention across the multimodal sequence. The lightweight configuration uses one transformer block, two attention heads, a 64-dimensional embedding, and a depthwise-convolutional feed-forward module. Squeeze-based recalibra-

tion further adjusts the contribution of informative global representations.

The convolutional representation $\mathbf{f}_{s,t}^c$, temporal HGRU representation $\mathbf{f}_{s,t}^h$, and global transformer representation $\mathbf{f}_{s,t}^v$ are recalibrated before adaptive fusion. Their fused representation is defined as

$$\boldsymbol{\eta}_{s,t} = \lambda_c \hat{\mathbf{f}}_{s,t}^c + \lambda_h \hat{\mathbf{f}}_{s,t}^h + \lambda_v \hat{\mathbf{f}}_{s,t}^v, \quad (5)$$

where, $\hat{\mathbf{f}}_{s,t}^c$, $\hat{\mathbf{f}}_{s,t}^h$, and $\hat{\mathbf{f}}_{s,t}^v$ denote the recalibrated convolutional, HGRU, and Health Vision Transformer representations, respectively, and λ_c , λ_h , and λ_v denote their dynamically learned fusion weights. The resulting embedding $\boldsymbol{\eta}_{s,t}$ combines local variation, temporal deterioration, and global multimodal interactions in a patient-specific risk representation.

E. Federated Training and Risk Prediction

Federated training was performed across five simulated healthcare-site computing clients. Each client represents a computing node that stores its own mutually exclusive subset of patient data, executes local model training, and communicates model parameters rather than raw health records to the aggregation server. This configuration preserves patient-level data separation throughout federated optimization.

At communication round r , the locally trained model parameters are combined using weighted federated averaging:

$$\Theta^{r+1} = \sum_{k=1}^K \frac{|\mathcal{D}_k|}{|\mathcal{D}|} \Theta_k^{r,E}, \quad (6)$$

where, K is the number of participating client nodes, $|\mathcal{D}_k|$ is the number of training samples stored at client k , $|\mathcal{D}| = \sum_{k=1}^K |\mathcal{D}_k|$ is the total number of federated training samples, E denotes the number of local training epochs, and $\Theta_k^{r,E}$ represents the locally updated parameters of client k after round r . The aggregated global parameters Θ^{r+1} are redistributed to the participating client nodes for the next communication round.

The fused representation $\eta_{s,t}$ is passed to a fully connected prediction layer. Binary prediction distinguishes stable from upcoming exacerbation states, while the three-class task estimates low-, moderate-, and high-risk states. A weighted focal loss objective downweights the impact of class imbalance and upweights hard samples. L_2 regularization was used to limit model complexity.

Patient level partitioning was enforced during training, validation, and testing to ensure no participant bleeding between data splits. During federated experiments, training participants were distributed amongst five simulated healthcare-site workers wherein each computing node stores a non-overlapping subset of patients and performs local model optimization. Adam was used for optimization with a learning rate of 1×10^{-4} . Dropout and weight decay were used with early stopping based on validation performance.

F. Evaluation and Model Interpretation

Model performance was evaluated using accuracy, sensitivity, specificity, precision, F1-score, area under the receiver operating characteristic curve (AUC), and Cohen's kappa. Macro-averaged measures were used for the three-class task to summarize performance across low-, moderate-, and high-risk groups. Computational efficiency was assessed using the number of trainable parameters and inference time per sample.

Model interpretation examined seven health-information groups: respiratory function, physiological instability, activity reduction, sleep disturbance, environmental exposure, mobility decline, and clinical background. The resulting importance patterns summarize the relative model emphasis assigned to these health-information groups during COPD exacerbation-risk prediction and are interpreted as predictive associations rather than causal effects.

IV. RESULTS

All experiments were implemented in Python 3.10 using PyTorch 2.2 with CUDA acceleration on an Intel Core i9 workstation with 64 GB RAM and an NVIDIA RTX 4060 GPU with 32 GB memory. FedSqueeze-HGRUViTNet was trained using Adam with an initial learning rate of 1×10^{-4} , a batch size of 128, dropout regularization, and validation-based early stopping.

All data partitions were constructed at the patient level to prevent participant overlap between training, validation, and testing subsets. For TOLIFE, 1,080,000 monitoring windows were divided into 756,000 training, 108,000 validation, and 216,000 held-out testing windows. The harmonized MIMIC-IV COPD cohort contained 300,000 samples, comprising 210,000 training, 30,000 validation, and 60,000 held-out testing samples. MIMIC-IV was used as a complementary electronic

health record (EHR)-based evaluation cohort using clinically compatible respiratory, physiological, demographic, laboratory, and clinical feature groups.

Each experiment was repeated using five random seeds with fixed patient-level partitions. The reported mean $\pm$ standard deviation therefore quantifies training variability rather than between-patient uncertainty. Although TOLIFE generated 216,000 highly overlapping test windows, these originated from only 15 independent test participants. The receiver operating characteristic (ROC) curves and confusion matrices present a run on the fixed 20% held-out test partitions. Performance was assessed using accuracy, sensitivity, specificity, precision, F1-score, area under the ROC curve (AUC), Cohen's kappa, inference time, and parameter count.

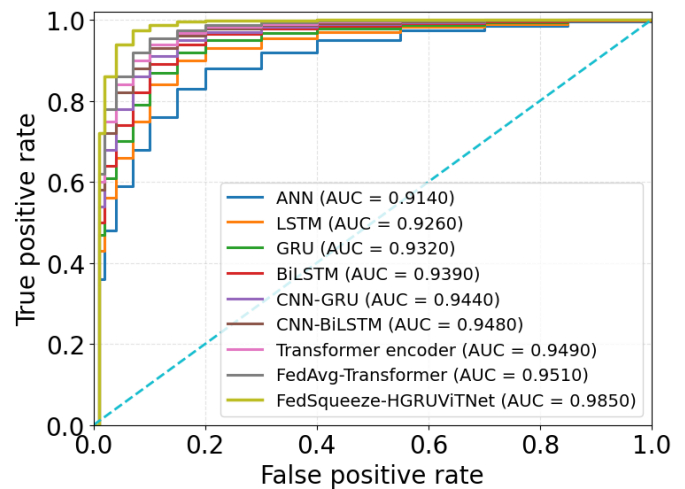


Fig. 2. ROC curves of deep learning and federated learning models on the TOLIFE held-out test partition.

Fig. 2 compares FedSqueeze-HGRUViTNet with artificial neural network (ANN), long short-term memory (LSTM), gated recurrent unit (GRU), bidirectional LSTM (BiLSTM), CNN-GRU, CNN-BiLSTM, Transformer Encoder, and federated averaging (FedAvg)-Transformer baselines. In the held-out run, FedSqueeze-HGRUViTNet achieved the highest AUC of 0.985, indicating stronger separation between stable and upcoming exacerbation samples.

Fig. 3 shows confusion matrices from held-out test partitions. TOLIFE's binary task had correctly classified 210,168 out of 216,000 test windows (accuracy of 97.30%) and correctly classified 208,915 windows for the three-class task (accuracy of 96.72%). For MIMIC-IV, our classifier correctly classified 57,972 of the 60,000 held-out samples, for 96.62% accuracy. We observe consistently high predictive performance between wearable health-index and EHR-based evaluation settings.

Table I summarizes performance across the five independent runs. FedSqueeze-HGRUViTNet achieved $97.30 \pm 0.28\%$ accuracy and an AUC of 0.985 ± 0.003 on TOLIFE. On the complementary MIMIC-IV EHR-based cohort, it achieved $96.62 \pm 0.39\%$ accuracy and an AUC of 0.978 ± 0.004 . The limited variation across runs indicates stable training behavior under both evaluation settings.

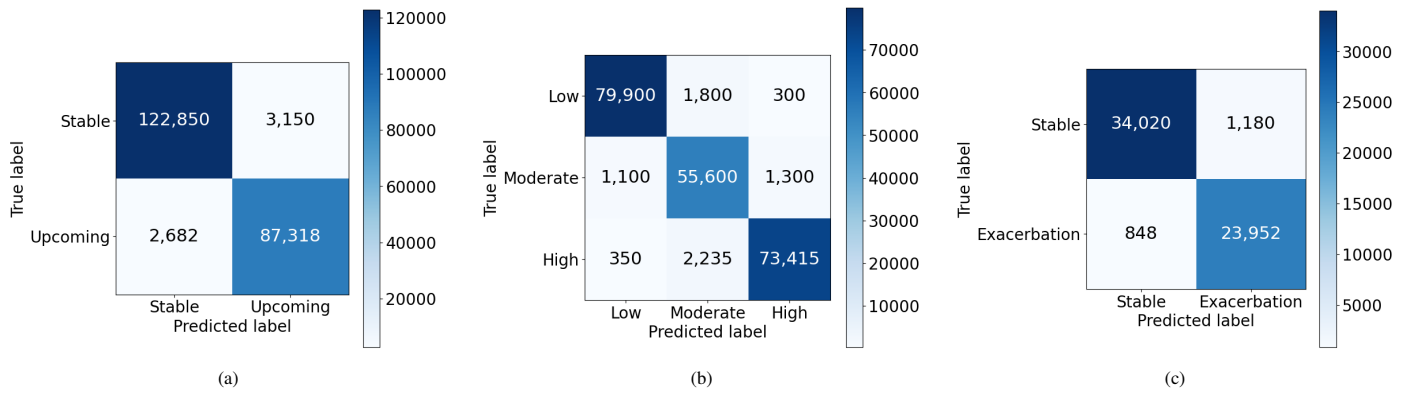


Fig. 3. Confusion matrices on the fixed 20% held-out test partitions: (a) TOLIFE binary classification of stable and upcoming exacerbation states; (b) TOLIFE three-class classification of low-, moderate-, and high-risk states; and (c) MIMIC-IV binary classification of stable and exacerbation-related admission samples.

TABLE I. TWO-CLASS EXACERBATION-RISK PREDICTION PERFORMANCE ACROSS FIVE RUNS

Dataset	Prediction task	Classification performance				Discrimination and agreement	
		Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-score (%)	AUC	Kappa
TOLIFE COPD Health Indices Dataset [26]	Stable vs. upcoming exacerbation	97.30±0.28	97.12±0.34	97.58±0.31	97.21±0.30	0.985±0.003	0.946±0.007
MIMIC-IV Clinical Database COPD cohort [27]	Stable vs. exacerbation-related admission	96.62±0.39	96.31±0.45	96.91±0.41	96.43±0.42	0.978±0.004	0.932±0.009

TABLE II. PERFORMANCE COMPARISON WITH COPD PREDICTION MODELS

Method	Binary accuracy (%)	Binary F1-score (%)	Three-class accuracy (%)	MIMIC-IV accuracy (%)	Parameters (% of ours)	Inference (ms)
XGBoost-SHAP [17]	89.46	89.18	88.71	88.25	52.14	2.91
DeepSpiro [11]	93.88	93.61	92.81	92.26	146.91	7.84
COPD-MMDDxNet [13]	94.56	94.31	93.36	92.84	184.72	9.37
COPD-TransNet [14]	95.02	94.75	93.84	93.31	168.09	8.64
CNN-BiLSTM-ViT [8]	95.18	94.96	94.02	93.56	203.19	10.44
Telemonitoring NN [23]	95.31	95.07	94.12	93.74	132.64	7.58
FedSqueeze-HGRUViTNet	97.30	97.21	96.72	96.62	100.00	5.18

The comparison architectures in Table II were adapted and evaluated using the same preprocessing, patient-level partitioning, and evaluation pipeline. FedSqueeze-HGRUViTNet resulted in the best performance across metrics with a binary accuracy of 97.30%, three-class accuracy of 96.72%, and MIMIC-IV accuracy of 96.62%. The best performing comparison method had a binary accuracy of 95.31% resulting in a gap of 1.99 percentage-points. FedSqueeze-HGRUViTNet took 5.18 ms per-sample on the RTX 4060gpu. While XGBoost-SHAP took 2.91 ms. These numbers reflect workstation-level inference times; we did not measure edge-device latency.

Fig. 4 visualizes how the feature organization shifts after adaptive fusion. We see more overlap in the original representations, specifically between moderate- and high-risk samples. Meanwhile, the representations that have undergone adaptive fusion cluster into tighter, distinct groups. We observe this trend across all evaluated COPD risk classes after adaptive fusion, suggesting the classes are better separated.

For federated experiments, we simulated five healthcare-

site computing clients. Each client held a disjoint partition of the patient-level monitoring records, performed local model training, and only shared model parameters with the aggregation server instead of raw health records. Weighted FedAvg was used to aggregate the local updates, then send the new global model back to clients.

Table III compares centralized, local-only, and federated training configurations. FedSqueeze-HGRUViTNet achieved the highest reported accuracy of $97.30 \pm 0.28\%$, whereas centralized Squeeze-HGRUViTNet achieved $92.86 \pm 0.41\%$. This difference describes the evaluated configurations but does not independently establish a benefit from federated optimization, since equivalent hyperparameter tuning and training budgets are not documented.

Fig. 5 compares the attention patterns generated by the evaluated models. FedSqueeze-HGRUViTNet produced more localized and class-distinct responses across the evaluated risk states than the comparison models, reflecting differences in the learned health-index representations.

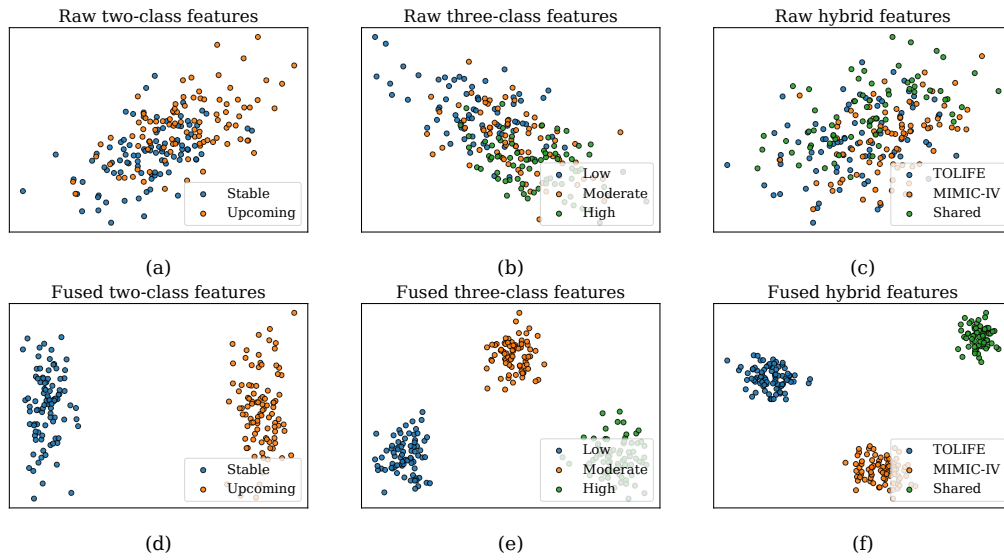


Fig. 4. T-distributed stochastic neighbor embedding (t-SNE) visualization of learned representations. Panels (a)–(c) show the original feature distributions for binary, three-class, and combined-dataset evaluations, while panels (d)–(f) show the corresponding representations after adaptive feature fusion.

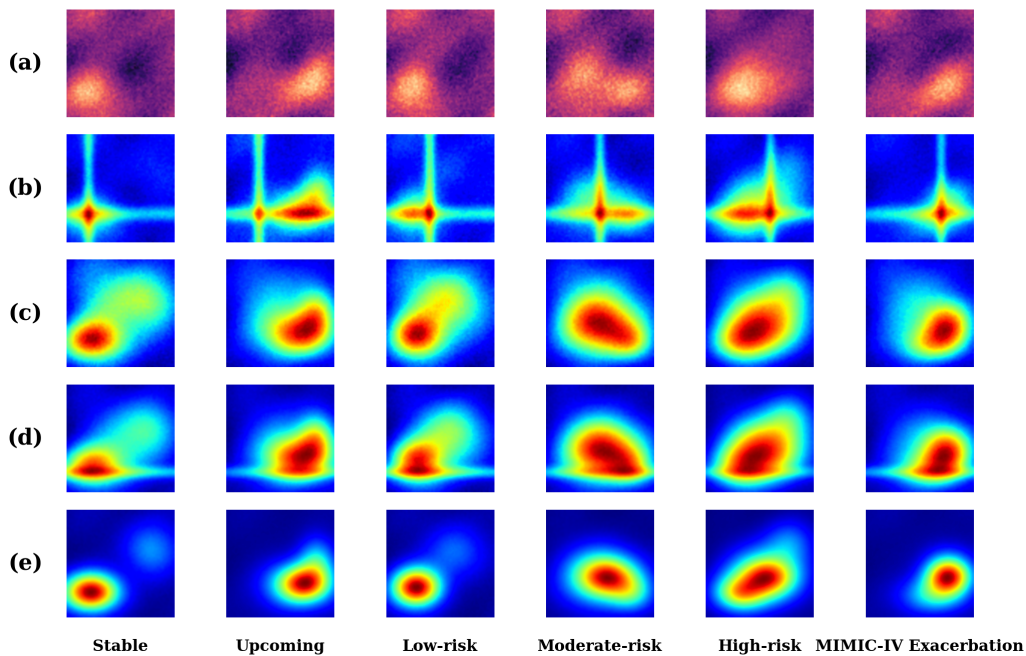


Fig. 5. Attention-map visualization of COPD health-index representations for the original input, CNN-BiLSTM, Transformer Encoder, FedAvg-Transformer, and FedSqueeze-HGRUViTNet. Columns represent stable, upcoming exacerbation, low-risk, moderate-risk, high-risk, and MIMIC-IV exacerbation cases.

Table IV summarizes performance under data and federated-distribution perturbations. Accuracy remained at or above 95.62% under 30% feature missingness, 20% Gaussian-noise perturbation, and high non-independent and non-identically distributed (non-IID) client heterogeneity. The 14-day monitoring window achieved the highest accuracy, while the 7-day and 21-day configurations reduced accuracy by 1.42 and 0.18 percentage points, respectively.

Fig. 6 shows the federated optimization behavior and a patient-level risk trajectory. As shown in Fig. 6a, global accuracy increased rapidly during the initial communication

rounds and stabilized after approximately 18–20 rounds, while validation loss and client drift progressively decreased.

In the 14-day sequence shown in Fig. 6b, the predicted risk remained near the stable baseline during the initial monitoring period, increased during the designated exacerbation-risk interval, and decreased during the subsequent post-risk monitoring period. Together with the robustness results, these observations characterize model behavior under changes in monitoring quality and federated client distributions.

TABLE III. CENTRALIZED AND FEDERATED LEARNING COMPARISON FOR TWO-CLASS EXACERBATION-RISK PREDICTION

Training setting	Data location	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-score (%)	AUC
Centralized Squeeze-HGRUViTNet	Central server	92.86±0.41	92.54±0.47	93.21±0.44	92.68±0.45	0.941±0.006
Local-only Squeeze-HGRUViTNet	Client nodes	93.84±0.61	93.25±0.68	94.38±0.63	93.52±0.65	0.953±0.007
FedAvg-GRU	Client nodes	95.08±0.46	94.71±0.52	95.51±0.48	94.83±0.49	0.947±0.006
FedAvg-Transformer	Client nodes	95.31±0.42	94.96±0.49	95.72±0.45	95.07±0.46	0.951±0.006
FedSqueeze-HGRUViTNet	Client nodes	97.30±0.28	97.12±0.34	97.58±0.31	97.21±0.30	0.985±0.003

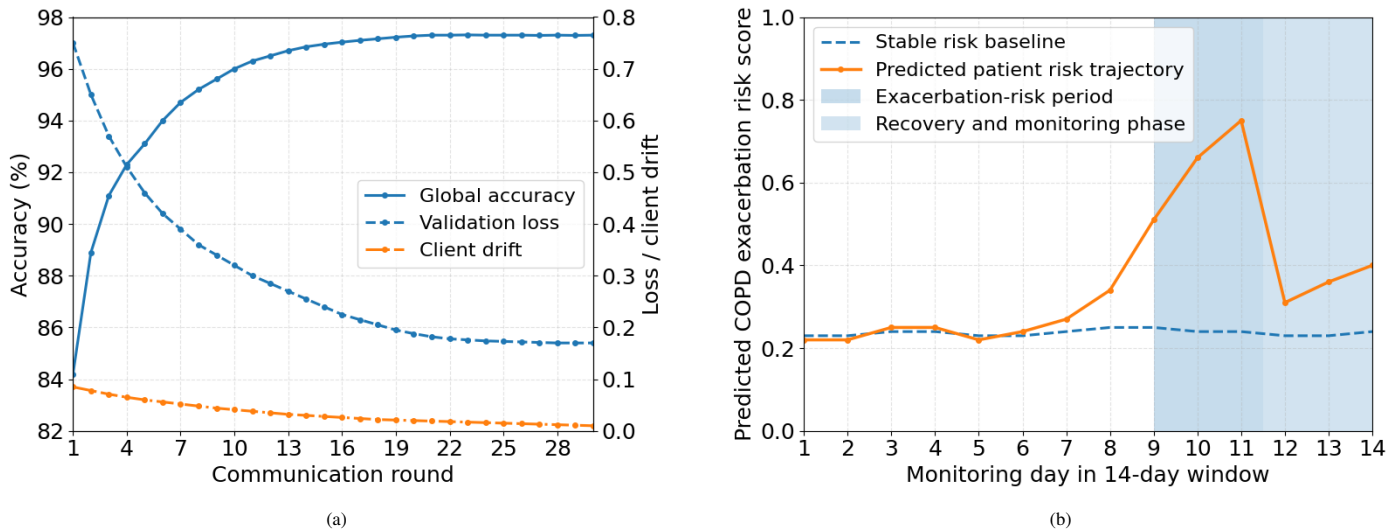


Fig. 6. Federated training and patient-level risk behavior: (a) global accuracy, validation loss, and client drift across communication rounds; and (b) predicted exacerbation-risk trajectory for a held-out monitoring sequence.

TABLE IV. ROBUSTNESS OF FEDSQUEEZE-HGRUViTNET UNDER PERTURBATIONS

Condition	Setting	Accuracy (%)	AUC	Drop (%)
Missingness	30% missing	95.62±0.51	0.966±0.006	1.68
Noise	20% Gaussian	95.74±0.55	0.968±0.006	1.56
Window length	7 days	95.88±0.47	0.970±0.006	1.42
Window length	14 days	97.30±0.28	0.985±0.003	0.00
Window length	21 days	97.12±0.31	0.983±0.004	0.18
Non-IID distribution	Mild	96.92±0.34	0.981±0.004	0.38
Non-IID distribution	High	95.81±0.56	0.969±0.006	1.49

V. DISCUSSION

FedSqueeze-HGRUViTNet achieved strong COPD exacerbation-risk prediction performance across the evaluated wearable and EHR-based settings. On TOLIFE, the model achieved 97.30% binary accuracy with an AUC of 0.985 and 96.72% three-class accuracy. The complementary MIMIC-IV evaluation yielded 96.62% accuracy and an AUC of 0.978. These findings indicate that combining local health-index variation, patient-specific temporal progression, and broader multimodal interactions provides a discriminative representation of COPD exacerbation risk.

Existing COPD models have used spirogram signals, pulmonary-function curves, computed tomography images, respiratory sounds, biomarkers, and hospital-based clinical variables [11], [12], [13], [14]. Respiratory-sound approaches

have also demonstrated the usefulness of convolutional, recurrent, and attention-based representations [8], [16], [19]. FedSqueeze-HGRUViTNet differs by jointly representing longitudinal respiratory, physiological, behavioral, sleep, mobility, environmental, and clinical information within a lightweight federated architecture.

Clinical-variable, biomarker, and telemonitoring approaches have also demonstrated useful COPD risk discrimination [17], [21], [23], [25]. Patient-specific baseline deviation and missingness info is captured alongside recurrent progression learning in FedSqueeze-HGRUViTNet. Age- and race-related effects that generalize across clients are learned through parameter-efficient Health Vision Transformer. These methods therefore learn complementary local, temporal, and global representations through convolutional, HGRU, and transformer branches respectively; combined through adaptive fusion.

Under missing observations, Gaussian-noise perturbation, monitoring-window shifts, and high non IID client heterogeneity; model performance degraded gracefully. Federated optimization also converged past 18–20 communication rounds. All tests concluded with maintained accuracy (95.81%) under the high-heterogeneity condition. Care should be taken not to over-interpret these results beyond this specific experimental setup.

MIMIC-IV provides complementary evaluation of retro-

spective COPD-related admission classification using hospital-based EHR features, whereas TOLIFE evaluates upcoming-exacerbation risk using longitudinal wearable information. The reported 96.62% MIMIC-IV accuracy should therefore not be interpreted as evidence of prospective exacerbation forecasting or cross-dataset parameter transfer, neither of which is established by the reported experiments.

The main limitation is the small number of independent TOLIFE participants (74, including 15 held-out patients) despite the large number of overlapping monitoring windows. Fixed-partition seed variation does not establish patient-level confidence intervals. Furthermore, the five healthcare clients were simulated, and communication overhead and privacy leakage were not evaluated. Clinical alert burden, actionable prediction lead time, and decision utility remain unquantified. The t-SNE and attention visualizations illustrate representation patterns but do not establish explanation faithfulness. Larger participant-level evaluations, component ablations, and prospective clinical validation are needed.

VI. CONCLUSIONS

This study presented FedSqueeze-HGRUViTNet, a lightweight federated deep learning model for COPD exacerbation-risk prediction from multimodal health information. The model combines depthwise separable convolution, health-gated residual GRU learning, squeeze-enhanced Health Vision Transformer attention, adaptive feature fusion, and federated optimization to capture local variation, patient-specific deterioration, and broader interactions across heterogeneous health indices. On the TOLIFE COPD Health Indices Dataset, the model achieved 97.30% accuracy for binary exacerbation-risk prediction and 96.72% accuracy for three-class risk stratification, while the complementary MIMIC-IV EHR-based evaluation achieved 96.62% accuracy. Across five independent runs, the model showed stable performance, and the robustness experiments demonstrated limited degradation under missing observations, Gaussian-noise perturbation, monitoring-window variation, and non-IID client distributions. These findings indicate that FedSqueeze-HGRUViTNet provides an effective and computationally efficient approach for distributed COPD risk prediction using wearable health indices and harmonized clinical information.

Future work will evaluate the model on larger, more diverse, and prospectively collected COPD cohorts across independently operating healthcare sites. Further studies will examine probability calibration, subgroup analysis, secure federated aggregation, personalized risk thresholds, and edge-supported deployment.

REFERENCES

- [1] P. Niraula, M. Upreti, S. Kadariya, B. Poudel, S. Kadariya, and S. Kunwar, *AI/ML-driven prediction of COPD exacerbations and readmissions: A systematic review and meta-analysis*, *Frontiers in Digital Health*, vol. 7, Art. no. 1641356, 2025.
- [2] C. Kapatais, T. Karaoulani, A. Papanikolaou, E. Koukaki, A. Bakakos, C. Anagnostopoulou, A. I. Papaioannou, P. Bakakos, and N. Rovina, *The role of artificial intelligence in predicting COPD exacerbations using multimodal data: A systematic review*, *ERJ Open Research*, vol. 12, no. 3, Art. no. 01420-2025, 2026.
- [3] L. A. Boyce and V. Prybutok, *Artificial intelligence for early detection and prediction of chronic obstructive pulmonary disease exacerbations*, *Healthcare*, vol. 14, no. 6, Art. no. 806, 2026.
- [4] M. Atzeni, G. Cappon, J. K. Quint, F. Kelly, B. Barratt, and M. Vettoretti, *A machine learning framework for short-term prediction of chronic obstructive pulmonary disease exacerbations using personal air quality monitors and lifestyle data*, *Scientific Reports*, vol. 15, Art. no. 2385, 2025.
- [5] E.-J. Ko, S.-O. Bae, and D. Kang, *A baseline study of interpretable machine learning using GC-MS breath VOCs for classifying asthma, bronchiectasis, and COPD*, *Scientific Reports*, vol. 15, Art. no. 44392, 2025.
- [6] H. Y. Park, S. Kong, M. Lee, H. Ryu, Y. Hamakawa, F. Luppi, and J. M. Leung, *Digital health technologies for improving the management of people with chronic obstructive pulmonary disease*, *Frontiers in Digital Health*, vol. 7, Art. no. 1640585, 2025.
- [7] R. Vázquez-Andrade, V. Alcaraz-Serrano, J. Buekers, P. Bufano, M. Laurino, A. Celi, H. Watz, J. Gea, N. Carbonaro, A. Tognetti, and J. Garcia-Aymerich, *Development of an artificial intelligence prediction model for moderate-to-severe COPD exacerbations using continuous multiple unobtrusive sensors: Protocol of a multicentre prospective observational study*, *BMJ Open Respiratory Research*, vol. 13, no. 1, Art. no. e003942, 2026.
- [8] X. Dong, X. Chen, Q. Chen, X. Liu, H. Chen, R. M. Aarts, and B. Yin, *A multi-scale hybrid efficient deep learning model for COPD detection using respiratory sounds*, *IEEE Journal of Biomedical and Health Informatics*, early access, 2026.
- [9] S. A. Marrah, J. Wang, A. B. Koroma, G. D. Kamara, O. S. Babatunde, M. Marrah, R. T. Stamber, and S. E. Saidu, *Federated deep learning for privacy-preserving disease detection in IoT-enabled healthcare systems*, *Frontiers in Computer Science*, vol. 8, Art. no. 1725597, 2026.
- [10] S. Panda, R. Dubey, B. Jena, V. Pareek, L.-W. Tsai, and S. Saxena, *Federated learning frameworks in smart e-healthcare: A systematic literature review with bias evaluation*, *Applied Soft Computing*, vol. 171, Art. no. 112747, 2025.
- [11] S. Mei, X. Li, Y. Zhou, J. Xu, Y. Zhang, Y. Wan, S. Cao, Q. Zhao, S. Geng, J. Xie, S. Chen, and S. Hong, *Deep learning for detecting and early predicting chronic obstructive pulmonary disease from spirogram time series*, *npj Systems Biology and Applications*, vol. 11, Art. no. 18, 2025.
- [12] E.-T. Jeon, H. Park, J.-K. Lee, E. Y. Heo, C. H. Lee, D. K. Kim, D. H. Kim, and H. W. Lee, *Deep learning-based chronic obstructive pulmonary disease exacerbation prediction using flow-volume and volume-time curve imaging: Retrospective cohort study*, *Journal of Medical Internet Research*, vol. 27, Art. no. e69785, 2025.
- [13] Y. Yi, L. Shi, H. Liu, M. Wang, M. Feng, and Y. Li, *COPD-MMDDxNet: A multimodal deep learning framework for accurate COPD diagnosis using electronic medical records*, *Frontiers in Medicine*, vol. 12, Art. no. 1601736, 2025.
- [14] A. Liu, B. Zhang, W. Li, M. Ao, X. Xiong, J. Mu, M. Deng, Q. Ye, X. Lu, and L. Yang, *COPD-TransNet: A Swin Transformer network with quantitative emphysema feature fusion for COPD detection and staging from opportunistic CT scans*, *Journal of Imaging Informatics in Medicine*, early access, 2026.
- [15] J. Zhou, J. Huang, Q. Wang, J. Yan, H. Cao, L. Huang, S. Chen, X. Ruan, W. Zhu, J. Mao, Y. Liu, Z. Bu, M. Yang, Q. Wang, Y. Zhou, E. Fan, L. Tong, X. Sun, D. Zhao, P. Wang, M. Zhou, and J. Qu, *A cough sound-based deep learning algorithm for accessible prompt detection of chronic obstructive pulmonary disease with smartphones*, *npj Primary Care Respiratory Medicine*, vol. 36, Art. no. 32, 2026.
- [16] T. Bikku, S. S. K. P. N. V., S. Thota, J. J. Pujari, R. K. Batchu, P. Mortezaagha, M. K. Kumar, and R. Anitha, *Deep learning-driven early diagnosis of respiratory diseases using CNN-RNN fusion on lung sound data*, *Scientific Reports*, vol. 15, Art. no. 45233, 2025.
- [17] D. Kuang, J. Min, H. Deng, Y. Zhao, Y. Sun, and J. Hong, *Application of machine learning models for predicting risk factors of acute exacerbations in chronic obstructive pulmonary disease*, *Frontiers in Medicine*, vol. 13, Art. no. 1804544, 2026.
- [18] Q. Wu, H. Guo, R. Li, J. Han, Z. Zhang, A. Jingsi, and S. Kang, *Multimodal data-driven multitask learning for enhanced identification and classification of chronic obstructive pulmonary disease: A retrospective study*, *Journal of Global Health*, vol. 16, Art. no. 04028, 2026.

- [19] G. Ayappan, S. Sumathi, V. Mani, and P. E. Jebarani, *Novel deep learning-based optimization framework for the classification of respiratory diseases using lung sound analysis*, Scientific Reports, vol. 16, Art. no. 19018, 2026.
- [20] Q. Wu, R. Li, H. Guo, J. Han, Z. Zhang, A. Jingsi, and S. Kang, *PulmoClass-3DAtt: A self-attention network for classification of COPD, PRISm, and normal*, International Journal of Chronic Obstructive Pulmonary Disease, vol. 21, pp. 1–13, 2026.
- [21] A. J. Yousuf, G. Parekh, M. Farrow, G. Ball, S. Graziadio, K. Wilson, C. Lendrem, L. Carr, L. Watson, S. Parker, J. Finch, S. Glover, V. Mistry, K. Porter, A. Duvoix, L. O'Brien, S. Rees, K. Lewis, P. Davis, and C. E. Brightling, *Artificial neural network risk prediction of COPD exacerbations using urine biomarkers*, ERJ Open Research, vol. 11, no. 3, Art. no. 00797-2024, 2025.
- [22] N. K. Rajpoot, P. D. Singh, and B. Pant, *A novel framework for COPD management in cyber-physical systems using machine learning*, Scientific Reports, vol. 15, Art. no. 36517, 2025.
- [23] J. Moraza, C. Esteban-Aizpiri, A. Aramburu, *et al.*, *Using machine learning to predict deterioration of symptoms in COPD patients within a telemonitoring program*, Scientific Reports, vol. 15, Art. no. 7064, 2025.
- [24] J. López-Canay, M. Casal-Guisande, A. Pinheira, R. Golpe, A. Comesaña-Campos, A. Fernández-García, C. Represas-Represas, and A. Fernández-Villar, *Predicting COPD readmission: An intelligent clinical decision support system*, Diagnostics, vol. 15, no. 3, Art. no. 318, 2025.
- [25] Y. Zhu, M. Wang, X.-N. Gu, C. Wang, and S.-M. Deng, *Development and validation of the machine learning model for acute exacerbation of chronic obstructive pulmonary disease prediction based on inflammatory biomarkers*, Frontiers in Medicine, vol. 12, Art. no. 1616712, 2025.
- [26] TOLIFE Project, *Health indices dataset from Clinical Study A – TOLIFE Project*, Zenodo, 2025. doi: 10.5281/zenodo.16642439
- [27] A. Johnson, L. Bulgarelli, T. Pollard, B. Gow, B. Moody, S. Horng, L. A. Celi, and R. Mark, *MIMIC-IV*, PhysioNet, version 3.1, 2024. doi: 10.13026/kpb9-mt58