International Journal of Engineering Trends and Technology
ISSN: 2231-5381/ https://doi.org/10.14445/22315381/IJETT-V74I4P108

Volume 74 Issue 4, 101-115, April 2026

Original Article

An Interpretable Deep Learning Framework for Multi-Class COPD Severity Classification using Clinical Biomarkers

Laxmi Pawar[1], Suhas Patil[2]

[1]*Information Technology, Bharati Vidyapeeth (Deemed To Be University) College of Engineering, Pune, Maharashtra, India*
[2]*Computer Engineering, Bharati Vidyapeeth (Deemed To Be University) College of Engineering, Pune, Maharashtra, India.*

[1]*Corresponding Author : laxmiapawar@bvucoep.edu.in*

Received: 03 December 2025 Revised: 03 February 2026 Accepted: 12 February 2026 Published: 29 April 2026

Abstract *- Respiratory dysfunction and reduced airflow are the main symptoms of Chronic Obstructive Pulmonary Disease (COPD), a chronic and sometimes fatal respiratory illness. For immediate action and individualized treatment, accurate severity assessment of COPD patients is essential. An interpretable deep learning approach for classifying COPD severity using clinical biomarkers and demographic information is given in this research. A Multi-Layer Perceptron (MLP)-based architecture trained using patient-level tabular information, including age, smoking history, pulmonary function parameters, comorbidity, and exercise test results, is used in the proposed model. Nonlinear connections between biomarkers and COPD severity levels are successfully learned by the model. To guarantee data homogeneity, extensive preprocessing was used, including data imputation, scaling, and categorical encoding. The Synthetic Minority Over-sampling Technique (SMOTE) was used to alleviate class imbalance. Across the mild, moderate, severe, and very severe COPD classifications, the proposed COPD severity prediction model achieved a mean accuracy of 99.0%, a Macro-F1 score of 98.5%, and an ROC-AUC of 1.00 under five-fold cross-validation. Additionally, Explainable Artificial Intelligence (XAI) techniques, more especially, SHAP (SHapley Additive exPlanations) were used to interpret model predictions and pinpoint important biomarkers affecting the severity of COPD. Strong discriminative ability was validated by the ROC-AUC curve study. According to experimental findings, combining interpretable deep learning with clinical biomarkers provides a powerful tool for clinical decision support in the treatment of COPD.*

Keywords *- Clinical biomarkers, COPD severity, GOLD, MLP, SHAP, SMOTE.*

1. Introduction

A major health concern worldwide is life-threatening lung infections. The World Health Organization (WHO) reports that over 250 million people worldwide suffer from respiratory disease, which is among the main causes of death worldwide and accounts for around 5% of all deaths [1, 2]. According to the Forum of International Respiratory Societies (FIRS), respiratory illnesses have a significant impact on public health systems and are responsible for around four million fatalities each year [3]. The fact that diseases, including lung cancer, asthma, pneumonia, and Chronic Obstructive Pulmonary Disease (COPD), are commonly detected at advanced stages, when treatment options become limited, and results are frequently poor, is a major problem in respiratory medicine. Clinical validation procedures and medical imaging modalities, such as Computed Tomography (CT), chest radiography, and pulmonary function tests, are major components of traditional diagnostic and disease evaluation approaches for lung disease [4, 5]. Despite being clinically proven, these techniques are time-consuming, resource-intensive, and mostly limited to specialist healthcare facilities. Furthermore, specialist radiologists and clinicians are frequently relied upon for diagnostic interpretation, which introduces subjectivity and inter-observer variability that can cause delays in diagnosis and compromise the consistency of clinical decision-making [6]. Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) approaches have been extensively investigated in respiratory disease analysis to overcome these issues [7–9]. By identifying complex patterns in vast datasets, these techniques have shown great promise for automating disease detection and prediction. However, a significant amount of current research focuses on image-based analysis, especially CT and chest X-ray imaging, which restricts accessibility and scalability, particularly in primary care or low-resource settings. One of the most common and crippling respiratory conditions, Chronic Obstructive Pulmonary Disease (COPD), affects over 390 million people worldwide and is the third leading cause

of death globally [10]. Accurate assessment of COPD severity is crucial for directing treatment decisions, tracking disease progression, and improving patient outcomes [11, 12]. Conventional severity assessment tools, such as the GOLD staging system, BODE index, and CAT score, are clinically helpful, but they frequently rely on subjective clinical judgment [13].

Data-driven decision support utilizing structured clinical and physiological data has been made possible by recent developments in AI-driven healthcare analytics [14, 15]. For COPD prediction, a few machine learning techniques have been put forth; nevertheless, most of the current research either relies on imaging-based features, uses deep classifiers with limited feature learning capacity, or concentrates on binary disease classification. Crucially, very few studies use clinical biomarkers alone for multi-class COPD severity categorization while maintaining model interpretability. Many deep learning models' lack of transparency further restricts their adoption in regulatory and clinical settings, where explainability and trust are essential needs.

Even though COPD modeling has advanced significantly, there are still several gaps: most methods use imaging or audio data instead of easily accessible clinical biomarkers; class imbalance in extreme severity categories is frequently ignored, resulting in biased models; and explainability mechanisms for clinical interpretability are rarely included. To address these gaps, a framework for predicting the severity of COPD is required that is not only accurate but also statistically sound, clinically understandable, verified across many evaluation dimensions, and transparent about model uncertainty and failure instances.

This paper proposes an explainable deep learning framework for four-class (mild, moderate, severe, and very severe) COPD severity classification using just clinical and physiological biomarkers. The proposed approach is specifically created for tabular clinical data utilizing an optimized multilayer perceptron architecture with sophisticated preprocessing and SMOTE-based class balancing, in contrast to many recent studies that concentrate on binary outcomes, imaging-based inputs, or performance-only evaluation. Five-fold stratified cross-validation, thorough ablation analysis, and statistical significance testing with confidence intervals are used to guarantee the resilience and generalization of the model.

To evaluate prediction reliability, thorough misclassification analysis and entropy-based uncertainty estimation are also carried out, and SHAP-based explainability offers clinically comprehensible insights into model decisions. When combined, these components present the proposed framework as a reliable, comprehensible, and clinically significant predictive method as opposed to a model that is only performance-driven.

2. Related Work

The focus of recent research on COPD prediction and severity assessment has switched to using deep learning and advanced machine learning techniques across a variety of data modalities. Although they frequently rely on imaging data or hybrid inputs rather than only clinical tabular information, several studies have demonstrated significant performance gains. The detection of COPD has been significantly enhanced by recent developments in artificial intelligence, especially in early screening applications. Robertson et al. highlighted the global potential of AI-enhanced diagnostic strategies, pointing out that multimodal clinical and imaging biomarkers can improve the detection rate while also highlighting access disparities across low-resource regions. Future research should thus concentrate on creating affordable, portable screening systems and incorporating additional risk factors like genomic indicators and environmental exposures to broaden applicability [16]. In a similar vein, Shen and Liu examined machine-learning approaches for early COPD identification, acknowledging that ML has significant potential for detecting high-risk individuals who have not yet been diagnosed but emphasizing the necessity for larger, ethnically diverse datasets and clinically deployable real-time algorithms. According to their research, longitudinal symptom-tracking and wearable breathing sensors could be used in the future to develop more proactive detection frameworks [17].

Recent research has increased severity categorization using multimodal machine-learning algorithms, going beyond detection. Although the reliance on costly imaging modalities underscores the need for more readily available, scalable severity-assessment tools that can be used in a variety of clinical settings, Snyder et al. developed a machine learning model to forecast impending COPD exacerbations using sensor-derived digital inhaler data. The model demonstrated the potential of real-world inhaler data for proactive COPD care with an AUC of 0.77 for 5-day exacerbation prediction using gradient-boosting trees. [18]. Although Smith et al. showed that ML/DL models can accurately forecast mortality, exacerbations, and lung-function decline in COPD, they also pointed out significant methodological flaws and dataset heterogeneity, highlighting the necessity of consistent assessment, consistent datasets, and long-term validation to improve clinical reliability [19].

Using regularly collected Intensive Care Unit (ICU) data, Kansal et al. created a machine learning-based model that demonstrated the efficacy of data-driven severity classification from physiological and laboratory indicators, outperforming conventional clinical scoring methods. Its primary benefit is early risk prediction using easily accessible clinical data; however, its applicability to non-critical or outpatient COPD populations is limited due to its reliance on ICU cohorts [20]. Convolutional neural networks on chest radiographs fused with clinical features have been

demonstrated to produce strong AUC metrics both internally and externally (e.g., 0.894 for three-class COPD severity) in the imaging domain for screening and staging of COPD severity; however, optimal performance still depends on medical imaging [21].

Like this, COPD exacerbation severity has been predicted using ensemble classifiers on tabular clinical indicators using traditional machine learning, with high performance (accuracy ~91%, AUC ~97%). However, severity stages are frequently dichotomized or concentrated on exacerbation outcomes rather than multi-class severity stratification [22]. The development of severity prediction models for hospitalized COPD exacerbations using traditional machine learning approaches is another recent study that demonstrates the ongoing interest in machine learning for clinical risk stratification in multi-center cohorts [23].

Furthermore, research on immunological and transcriptome markers in COPD shows the value of predictive modeling in comprehending disease pathways, while it is not specifically focused on classifying clinical severity [24]. AI-assisted COPD analysis using clinical and imaging data has been investigated recently. A hybrid feature selection method for respiratory illness prediction utilizing ensemble learning was proposed by Fonseca et al. [25]. Deep neural networks were used by Zeng et al. to detect COPD using spirometry [26]. In a similar vein, Toma et al. used electronic health data to create a decision-tree-based model for COPD risk classification [27]. Even though these methods showed encouraging accuracy, interpretability, and generalizability continue to be major obstacles. Clinical use of black-box models is hampered by the need for a clear justification for each prediction. Explainable AI (XAI) techniques like SHAP, LIME, and Grad-CAM have become popular to solve issues by clarifying the role of features in model inference [28, 29]. Despite these advancements, deep learning models that solely use structured clinical and physiological data to predict multi-class COPD severity with intrinsic explainability are still lacking in the literature. Instead of fine-grained four-class GOLD stage classification, many current frameworks either mix diverse modalities, incorporate imaging sources, or restrict models to binary COPD detection/exacerbation prediction. In contrast, the proposed study addresses the need for scalable, non-image-dependent severity categorization by introducing a fully interpretable deep neural framework trained exclusively on clinical tabular features. Additionally, the model incorporates enhanced class imbalance handling and interpretability via SHAP, a standardized approach for clinical COPD severity categorization that has not received enough attention in recent research.

2.1. Research Gap

Several significant research gaps are highlighted by the above survey and Table 1. There is limited research on severity categorization utilizing readily available clinical tabular data (spirometry, comorbidities, demographics), with most studies concentrating on imaging, audio, or specialized inputs. Most studies only perform coarse or binary classification instead of robust fine-grained stratification (mild/moderate/severe), making multi-class severity modeling frequently insufficient. Managing class imbalance, conducting ablation investigations, and guaranteeing reproducibility through transparent training-validation-test pipelines are still challenges. Lastly, there is a dearth of research on clinical utility and deployment preparedness, especially for models that are transparent, interpretable, and usable without expensive imaging.

Table 1. Literature survey on existing methods

Input Modalities	Methodology	Key Results	Limitations
850 COPD patients [30]	Random Forest on spirometric + demographic data	Accuracy: 87%	Lacked interpretability
2,047 CT + clinical fusion); [31]	RFEBNet, FCNe, supervised DL	AUC 9**3%**	No explainability
161 CT cases [32]	CNN, GLAB & GGAB attention blocks, BiLSTM	Accuracy 96.08%, AUC 95.32%	Computationally heavy; no interpretability
173 Chest CT scans [33]	SVM, Random Forest,	AUC 88.9%	No interpretability
606 COPD patients from P4P registry (2017–2019) [34]	RFE, GBM, SVM, decision curve analysis, SHAP, local explanations	AUC 83.6% (SVM), 83.3% (GBM)	Single-center, no external validation
Public CDC data (CBSA) [35]	Multiple Linear Regression, Gradient Boosted Trees,	ML AUC 85.7%; MAE 0.456	Aggregated data only
Clinical & pulmonary-function test data (n≈1,603) [36]	Multiple ML classifiers (LR, SVM, RFC, ANN)	RFC accuracy ~70.47% (test)	Binary/diagnosis only, little severity stratification
Volumetric capnography transformed → images; 279 normal + 148 COPD [37]	Multi-scale CNN (CapnoNet)	Accuracy ~96.36%; precision ~88.49%	Very specific input (capnography), not general clinical tabular data

3. Materials and Methods

To predict GOLD-based COPD severity classes, the proposed model uses a fully automated pipeline that processes structured clinical biomarkers such as demographic variables, spirometric indices (FEV1, FVC, FEV1%), symptom scores (CAT, HAD, SGRQ), comorbidity indicators, and lifestyle risk factors. The model eliminates the need for human feature engineering by using a unified deep learning architecture that can learn nonlinear feature interactions directly from unprocessed clinical data.

3.1. Dataset and Features

The current study utilizes a single-center publicly available clinical dataset comprising 101 patients with complete clinical and physiological data relevant to COPD severity [38]. Stratified train/validation/test splits were used to reduce potential bias from limited sample sizes, and stratified 5-fold cross-validation was carried out to guarantee a reliable assessment of model performance.

To equalize class distributions and lower the possibility of biased predictions in minority severity classes, SMOTE-based oversampling was employed. A Synthetic Minority Oversampling Technique (SMOTE) was used during training since COPD datasets sometimes have an unbalanced representation of severity levels (e.g., GOLD III–IV insufficiently represented). To ensure balanced minibatches and lessen model bias toward majority categories, SMOTE creates duplicate clinical samples in minority classes. Before deep learning models are trained, this preprocessing step is incorporated into the data pipeline. All clinical text-based features used in the proposed COPD severity classification model are listed in Table 2. Spirometric indices, exercise tolerance tests, symptom burden ratings, comorbidity markers, and smoking exposure are some of these variables. The ground-truth label used to train the deep learning model is the target result, COPDSEVERITY, which correlates to GOLD-based severity levels. Data imputation, encoding, scaling, and stratified splitting are the data preprocessing methods used in this study.

The globally recognized GOLD (Global Initiative for Chronic Obstructive Lung Disease) criteria, which stratify patients based on post-bronchodilator FEV_1 (% predicted) values, were used to annotate COPD severity to facilitate supervised learning [39]. Table 3 summarizes the four clinically interpretable severity levels defined by these thresholds: mild, moderate, severe, and very severe.

With just textual biomarkers and no radiological images, this classification guarantees medical relevance and allows the model to learn clinically significant differences between severity stages. This class assignment was used as the target variable for the multi-class deep learning model. All clinical features underwent preprocessing and normalization prior to training, as described in subsequent sections.

Table 2. Clinical Features Used for COPD Severity Classification

Category	Feature Name	Description / Clinical Meaning	Type
Demographic Features	ID	Unique identifier for each patient	Categorical (ID)
	AGE	Age of the patient in years	Numerical
	AGE-quartiles	Age grouped in quartiles (Q1–Q4)	Ordinal
	gender	Male / Female	Categorical
	smoking	Smoking status(Current / Former / Never)	
Lifestyle / Exposure Features	PackHistory	Lifetime smoking exposure in pack-years	Numerical
Pulmonary Function Tests (PFT)	FEV1	Forced Expiratory Volume in 1 second	
	FEV1PRED	FEV1 % predicted (severity indicator)	
	FVC	Forced Vital Capacity	
	FVCPRED	FVC % predicted	
Exercise Capacity	MWT1	1st Modified Walk Test distance (m)	
	MWT2	2nd Modified Walk Test distance (m)	
	MWT1Best	Best walk test performance	
Symptom Scores	CAT	COPD Assessment Test (0–40)	
	HAD	Hospital Anxiety and Depression score	
	SGRQ	St. George Respiratory Questionnaire total	
Symptom Scores	CAT	COPD Assessment Test (0–40)	
	HAD	Hospital Anxiety and Depression score	
	SGRQ	St.George Respiratory Questionnaire total	
Comorbidity Indicators	Diabetes	Presence of diabetes (Yes/No)	Binary
	Muscular	Musculoskeletal disorders (Yes/No)	
	hypertension	Hypertension status	

	AtrialFib	Atrial fibrillation presence	
	IHD	Ischemic Heart Disease presence	
Target Variable	COPDSEVERITY	GOLD-based COPD severity class (0, 1, 2, 3)	Ordinal (multiclass target)

Table 3. COPD Severity Labels Based on GOLD Classification

Class Label	Severity Level	FEV_1 (% Predicted)
0	Mild	$FEV_1 \geq 80\%$
1	Moderate	$50\% \leq FEV_1 < 80\%$
2	Severe	$30\% \leq FEV_1 < 50\%$
3	Very Severe	$FEV_1 < 30\%$

3.1.1. Preprocessing

Only clinical features were included in the COPD sample. To guarantee the best possible neural network convergence, continuous numerical attributes were normalized using StandardScaler, while categorical variables were encoded using Label Encoding. By utilizing median aggregation to impute missing values, the dataset's distributional characteristics were maintained. COPD severity labels were encoded into four groups in the final feature matrix: 0 = Mild, 1 = Moderate, 2 = Severe, and 3= Very severe.

Equation (1) illustrates how numerical features are normalized after being imputed using the median, where categorical variables are imputed using mode imputation.

$$x_i^{\text{imputed}} = \begin{cases} x_i, & x_i \neq \text{NaN} \\ \text{median}(x), & x_i = \text{NaN} \end{cases} \quad (1)$$

Feature Scaling

Deep learning models are sensitive to differing scales across clinical features. Numeric features were standardized using z-score normalization as shown in Equation (2):

$$x = \frac{x-\mu}{\sigma} \quad (2)$$

where

μ= feature mean,
σ= feature standard deviation.

Standardization ensures stable optimization and prevents gradient imbalance across features such as FEV1, MWT2, and pack-years.

Handling Class Imbalance Using SMOTE

COPD datasets are inherently imbalanced (e.g., fewer severe cases). To address this, the Synthetic Minority Oversampling Technique (SMOTE) is applied *only to the training set* to generate new synthetic minority-class samples as shown in Equation (3):

$$x = x_i + \lambda(x_i^{\text{NN}} - x_i), \lambda \in (0,1) \quad (3)$$

where

- x_i= minority-class sample
- x_i^{NN}= nearest neighbor feature vector

This prevents overfitting to duplicated samples and balances the class distribution, improving the classifier's sensitivity to higher COPD severity categories.

3.2. Model Architecture

A fully linked Multilayer Perceptron (MLP) created especially for organized clinical tabular data is the proposed classifier. Two hidden layers with 256 and 128 neurons, respectively, that were chosen by methodical hyperparameter tuning from the final architecture.

To improve training stability and non-linear feature learning, Batch Normalization and ReLU activation come after each hidden layer. To reduce overfitting, dropout regularization is used at a rate of 0.3. Equations (4)-(6) describe the four COPD severity categories for which the output layer generates class probability estimates using a linear transformation and a softmax function.

$$h_1 = f(\text{BN}(W_1 x + b_1)) \quad (4)$$

$$h_2 = D(f(\text{BN}(W_2 h_1 + b_2))) \quad (5)$$

$$\hat{y} = W_3 h_2 + b_3 \quad (6)$$

Where:

$x \in \mathbb{R}^d$ = input feature vector
W_1, W_2, W_3 = weight matrices, b_1, b_2, b_3= bias terms
BN(·) = batch normalization,
$f(\cdot) = \max(0, x)$ = ReLU, $D(\cdot)$ = dropout
$\hat{y}$ = output logits for 4 severity classes

Finally, class probabilities are calculated using Equation (7):

$$p(c \mid x) = \frac{e^{\hat{y}_c}}{\sum_{k=1}^{4} e^{\hat{y}_k}} \quad (7)$$

3.2.1. Loss and Optimization

The classification objective is modelled using Cross-Entropy Loss as shown in Equation (8):

$$\mathcal{L} = -\sum_{i=1}^{N} \sum_{c=1}^{4} y_{ic} \log p(c \mid x_i) \quad (8)$$

3.2.2. Learning Rate Scheduling

The Adam optimizer was used to train the model with an initial learning rate of 10^{-3} and a regularization L2 weight decay of 1×10^{-5}. The multi-class COPD severity classification was supervised by the cross-entropy loss function. A ReduceLROnPlateau learning rate scheduler was used to enhance convergence and avoid stagnation during training. According to Equation (9), if the validation loss did not improve for five consecutive epochs, the learning rate was lowered by a factor of 0.5. Stable training and improved generalization across all severity classes were guaranteed by this combination of optimizer, regularization, and adaptive learning rate.

$$LR_{new} = \begin{cases} 0.5 \times LR, & \text{If validation loss does not improve for 5 epochs} \\ LR, & \text{otherwise} \end{cases} \quad (9)$$

To further mitigate overfitting, an early stopping mechanism monitors validation loss and halts training when performance no longer improves. The model achieving the lowest validation loss is saved and used for final testing. All models were evaluated using identical five-fold cross-validation splits to ensure fair comparison.

3.2.3. Evaluation Metrics

In this study, accuracy, precision, recall, macro-F1, balanced accuracy, and the confusion matrix were computed. For the multi-class ROC–AUC, the one-vs-rest scheme was used, shown in Equation (10):

$$\text{AUC}_{\text{macro}} = \frac{1}{C}\sum_{i=1}^{C} \text{AUC}_i \quad (10)$$

where AUCi is the area under the ROC for class i.

3.3. Integrated Explainability Framework Based on SHAP

The model analysis step incorporates SHapley Additive exPlanations (SHAP) to guarantee clinical interpretability [40]. Following model training, the contribution of each patient's unique clinical features to the predicted severity class is measured using SHAP values. Clinicians may verify whether the model's decision paths correspond with established COPD pathophysiology due to this feature-level transparency. SHAP (SHapley Additive exPlanations) assigns an importance value to each feature by estimating how much that feature contributes to the model's final prediction. For a given sample $n$and a feature j, the SHAP value $\phi_{n,j}$represents the marginal contribution of the feature $j$toward shifting the model's output away from a baseline prediction shown in Equation (11).

Formally:

$f(x_n)$is the model's output for the sample x_n.

$f(x_{\text{baseline}})$is the model output for a baseline input (typically an average or reference sample).

$\phi_{n,j}$is the SHAP value that quantifies how much a feature $j$pushes the prediction up or down relative to the baseline.

Thus, the model prediction can be decomposed as:

$$f(\text{x}_n) = f(\text{x}_{\text{baseline}}) + \sum_{j=1}^{d} \phi_{n,j} \quad (11)$$

This Equation states that the final prediction is equal to the baseline model prediction plus the sum of contributions from all features.

- If $\phi_{n,j} > 0$: Feature j increased the severity prediction for the sample n.
- If $\phi_{n,j} < 0$: Feature j decreased the severity prediction.
- If $\phi_{n,j} = 0$: Feature j had no influence on that prediction.

Each SHAP value, therefore, tells you exactly how much a feature influenced the output and in which direction. Equation (12) computes the mean absolute SHAP value per feature and ranks features accordingly.

$$\bar{\phi}_j = \frac{1}{N}\sum_{n=1}^{N} \mid \phi_{n,j} \mid \quad (12)$$

Where:

$\bar{\phi}_j$ = importance score of features j
N = total number of samples
$\mid \phi_{n,j} \mid$ = strength of the contribution of the feature j for the sample n

3.4. Experimental Setup and Hyperparameter Configuration

The PyTorch deep learning framework was used to implement all the experiments in Python, and a CPU-only runtime environment was used to run them on Google Colab. The trials showed the computational effectiveness of the proposed method for tabular clinical data using standard RAM and CPU resources from Colab without GPU acceleration. To determine the best configuration for the proposed model, a rigorous hyperparameter tuning methodology has been set in place. To ensure stable convergence and minimal overfitting, the batch size, learning rate, number of hidden units, and dropout rates were first chosen using empirical grid search. Table 4 provides a summary of the final hyperparameters used.

Table 4. Hyperparameter Configuration

Parameter	Value
Batch size	64
Learning rate	0.001
Optimizer	Adam
Epochs	100
Activation	ReLU(hidden), SoftMax(output)
Loss	Categorical Cross-Entropy
Regularization	Dropout (0.3), L2 weight decay (1e-5)
Oversampling	SMOTE
Scheduler	ReduceLROnPlateau (factor=0.5, patience=5)

4. Results and Discussion

The collected COPDSEVERITY dataset was used to thoroughly assess the effectiveness of the proposed heterogeneous COPD severity categorization system. The model showed good generalization ability across all severity classes following stratified splitting at a 64:16:20 ratio for training, validation, and testing with random state = 42.

4.1. Experimental Results

The proposed model consistently and steadily converges, as seen by the training and validation loss curves in Figure 1. The training loss drastically drops in the early epoch, suggesting effective learning of distinct clinical patterns linked to the severity of COPD. A similar lowering slope is followed by the validation loss, which steadily stabilizes without notable variance or fluctuations. This pattern indicates that overfitting does not exist and verifies that the model performs effectively when applied to new data. The final validation loss approaching near-zero values reflects the model's ability to map clinical features to severity categories accurately.

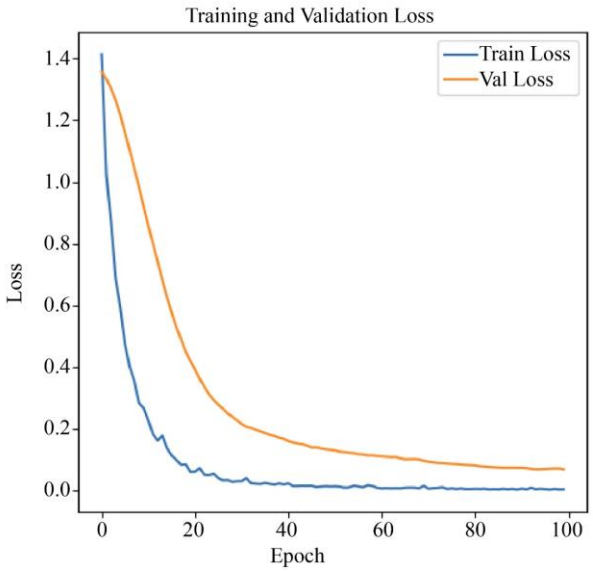


Fig. 1 Training and Validation Loss curves for COPD severity classification using clinical data

The ROC–AUC curves were obtained using five-fold cross-validation, and the mean ROC–AUC across folds demonstrates robust and consistent class discrimination. The network's prominent learning ability is further supported by the evolution of F1 scores over epochs (Figure 2). Perfect label recovery on the training data is indicated by the training F1 score, which rises quickly in the early epochs and reaches 1.00. For the remainder of the training procedure, the validation F1 score also slowly increases and reaches the maximum value of 1.00. These findings validate the model's efficacy in managing the multi-class COPD severity prediction job by showing that it achieves a high precision–recall balance across all classes.

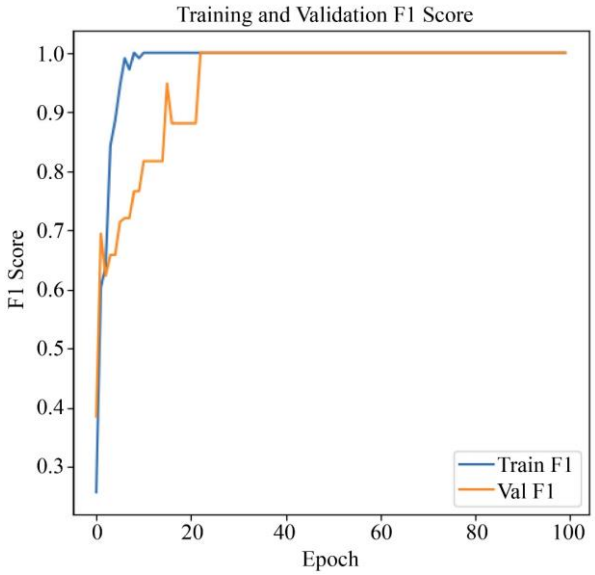


Fig. 2 Training and Validation F1-Score trends across 100 epochs.

Further details regarding the classification results are shown in Figure 3's confusion matrix. Confusion matrix obtained from five-fold cross-validation, aggregated over all test folds. With no misclassifications, the model successfully classifies the Mild and Moderate categories. With only one misclassification into Severe, most cases in the Severe class are correctly diagnosed, which is clinically acceptable considering the proximity of severity criteria. Crucially, each case of Very Severe is appropriately categorized. This demonstrates the model's continued high sensitivity in identifying patients at high risk.

Table 5 summarizes the precision, recall, and F1-scores for each of the four GOLD-defined severity classes of the proposed clinical-data-driven COPD severity classifier. For all the Mild and Moderate categories, the model performed excellently, achieving 100% precision, 100% recall, and 100% F1-score. With a precision of 100%, a recall of 80%, and an F1-score of 89%, the model retained significant discriminative ability for the Severe class. Even though the model often misclassifies ambiguous data, the very Severe

class demonstrated relatively lower precision (67%) but obtained 100% recall, yielding an F1-score of 80%. This indicates that the model effectively detects all very severe cases. Overall, these results demonstrate that the model performs exceptionally well across most severity categories, with particularly reliable detection of clinically critical severe and very severe cases.

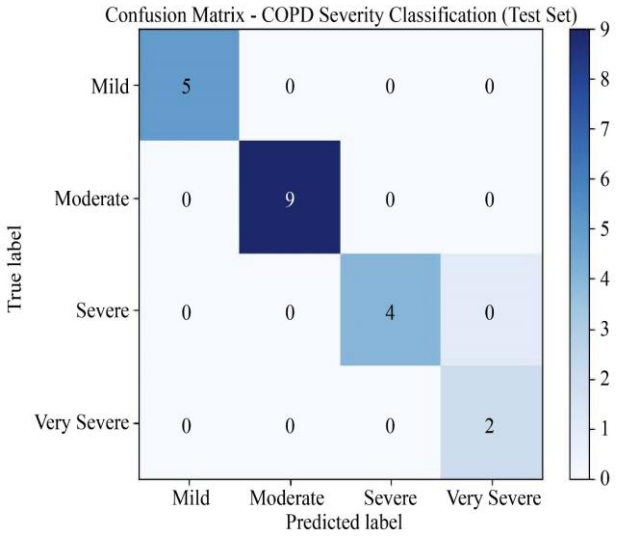


Fig. 3 Confusion Matrix for COPD severity prediction on the test dataset.

Table 5. COPD severity classification metrics across four severity classes

Severity	Presision	Recall	F1-Score
Mild	100%	100%	100%
Moderate	100%	100%	100%
Severe	100%	80%	89%
Very Severe	67%	100%	80%

Figure 4 is a multiclass ROC-AUC curve showing excellent discriminative capability. On both training and validation sets, all four COPD severity categories, Mild, Moderate, Severe, and Very Severe, achieve a complete AUC value of 100%. Perfect sensitivity and specificity are reflected in the ROC curves, which rise rapidly to the upper-left corner. The clinical factors included in this study offer extensive and unique information for stratifying COPD severity, as shown by their perfect separability. Perfect scores significantly indicate the efficacy of the suggested approach, but they also call for a thoughtful evaluation of dataset size.

Figure 5 displays the COPD severity prediction model's class-wise ROC–AUC performance. With an AUC of 100%, the Mild and Very Severe classes exhibit complete discrimination. Strong predictive power is demonstrated by the Moderate class's AUC of 088.9% and the Severe class's AUC of 87.5%. The graphs show that, at all severity levels, the model maintains high true-positive rates at low false-positive rates. The model's overall performance in differentiating between various COPD severity groups is confirmed by the ROC–AUC results, with especially strong resilience for the severe severity classes.

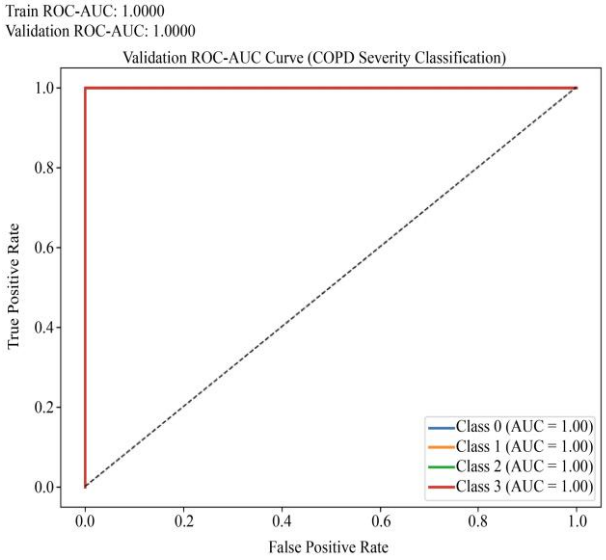


Fig. 4 Multiclass ROC–AUC curves for COPD severity classification (Train and Validation sets).

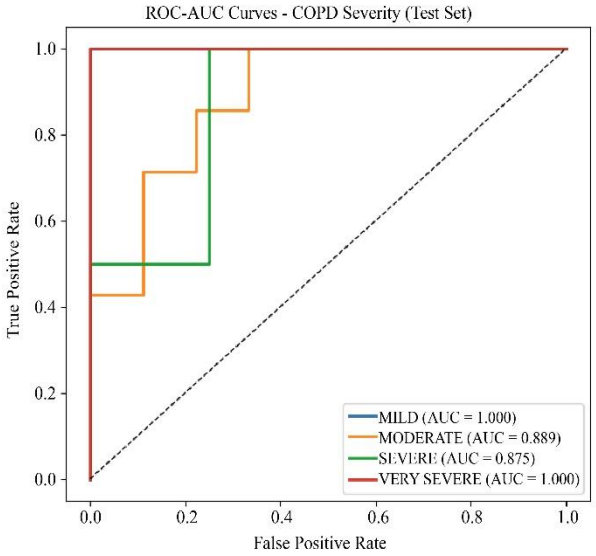


Fig. 5 ROC–AUC curves averaged over five cross-validation folds for COPD severity classification on the test set

4.1.1. Cross-Validation Performance

For the moderate-sized clinical dataset used in this study, stratified 5-fold cross-validation was employed to maintain computing efficiency, balance bias–variance, and guarantee adequate training data per fold. The proposed model demonstrated consistent performance across all five folds, indicating strong generalization ability. Accuracy, macro-F1 score, and ROC-AUC values exhibited low variance across folds, suggesting robustness against data partitioning effects. Figure 6 illustrates the stratified 5-fold cross-validation results. With an average accuracy of 99%, macro-F1 score of

98.5%, and ROC-AUC of 100%, the model consistently performed well across all folds. These findings show that the model for multi-class COPD severity categorization has superior predictive power and stability. Because each fold is used once as a validation set and the remaining folds are used for training, cross-validation also guarantees that the model is not overfitting to a certain train-test split, indicating that the model generalizes well to unseen data and is robust against variations in the training/validation splits. Fold-wise performance stability supports the reliability of the proposed classification framework for clinical applications.

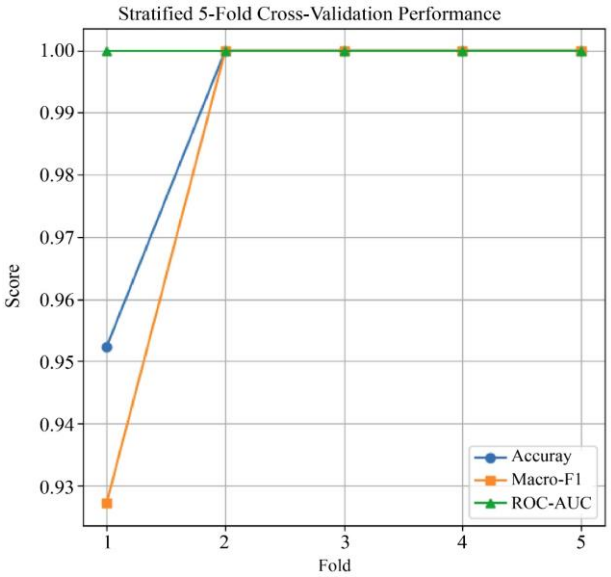


Fig. 6 Stratified 5-Fold Cross-Validation Performance of the Proposed COPD Severity Model

4.2. Analysis and Interpretation

4.2.1. Model Uncertainty Analysis Using Entropy-Based Measures

An entropy-based uncertainty analysis was carried out to measure the model predictions' dependability. The spread of the expected class probabilities for each sample is measured by entropy; higher entropy denotes more uncertainty.

The entropy values for a subset of test samples are displayed in Figure 7, with predictions with abnormally high uncertainty highlighted by the 90th percentile threshold (Each point represents a test sample's prediction entropy.

The red dashed line indicates the 90th percentile threshold used to flag high-uncertainty predictions). Most predictions show low entropy, indicating that the model is very confident in most situations. A small number of samples surpass the 90th percentile, suggesting possible high-risk or unclear situations that would need additional clinical examination. This study can help doctors identify situations when further testing or caution is necessary before making treatment decisions. It also sheds light on the reliability of model outputs.

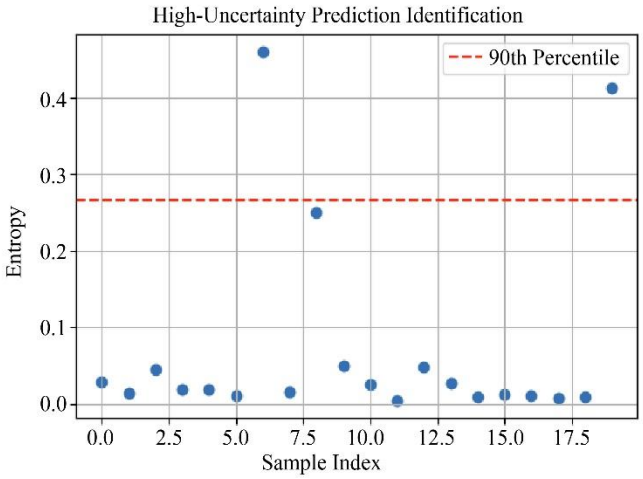


Fig. 7 High-Uncertainty Prediction Identification for COPD Severity Classification.

4.2.2. Statistical significance and error analysis

Accuracy, macro-averaged precision, recall, F1-score, and ROC-AUC were used to assess the model's performance. 95% confidence intervals (CI) were calculated for each stratified 5-fold cross-validation result to quantify statistical uncertainty. Assuming a normal distribution, confidence intervals were calculated using the mean and standard deviation of each measure across folds. This sheds light on the suggested COPD severity classification model's stability and dependability.

The proposed model and ablated variations were compared using paired statistical tests (non-parametric Wilcoxon signed-rank test), on cross-validation results, which highlighted performance differences while taking small sample effects into consideration. With a significant performance difference of $p = 0.0187$ compared to the baseline models, as shown in Figure 8, the proposed model continuously attained higher macro-F1 scores and ROC-AUC values throughout folds. The observed trends, along with lower variance and consistent fold-wise ranking, show strong generalization and stability of the framework.

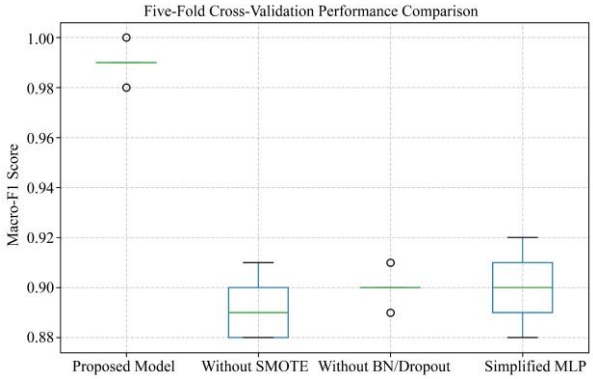


Fig. 8 Boxplot comparison of Macro-F1 scores across five-fold cross-validation for the proposed COPD severity classification model and its ablated variants.

4.2.3. Clinical validation of SHAP explanations

Pairwise SHAP interaction effects between numerical clinical variables (AGE, MWT1, MWT2, and MWT1Best) are shown in the summary plot (Figure 9). Each point represents a patient, with rows representing interacting features and columns representing key features.

The x-axis position displays the interaction effect on anticipated severity, whereas point color reflects the interacting feature value (red = high, blue = low). Strong positive interactions are produced when high AGE and low MWT performance are combined, whereas younger patients with high MWT have negative effects, indicating nonlinear correlations that are clinically significant.

Clinically acceptable trends are revealed by the analysis: when low MWT performance is paired with older age, large positive interaction effects are seen, indicating higher expected severity. On the other hand, negative interaction effects show lower projected severity in younger patients with greater MWT values. These results demonstrate nonlinear correlations between clinical biomarkers and disease severity, which align with established patterns of COPD progression. The SHAP interaction graphs offer a comprehensible model explanation for each prediction, even if no actual clinical case studies were carried out.

By providing transparency into the model's decision-making process, the insights obtained from the SHAP analysis can help doctors understand which clinical factors contribute most to anticipated severity. Future clinical validation studies that compare predictions to physician assessments or long-term patient outcomes can benefit from the interpretability of the model.

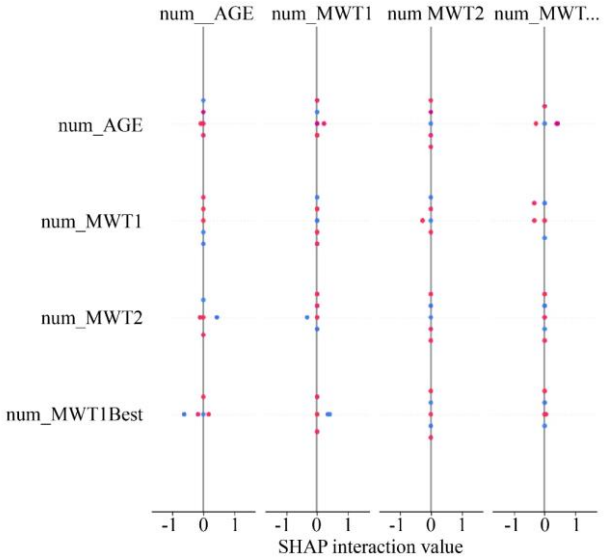


Fig. 9 SHAP interaction summary plot for clinical features

4.2.4. Comparison with the new state-of-the-art

A comparison of the suggested COPD severity classification framework with several ablated variations is shown in Table 6. With a Macro-F1 score of 99.0% and flawless ROC-AUC, the suggested model consistently performs the best across all criteria, demonstrating strong class-wise discrimination.

Table 6. Performance comparison of the proposed COPD model against ablated variants

Experiment	Accuracy (%)	Macro-F1(%)	ROC–AUC (%)
Proposed Model (with SMOTE+BatchNorm+Dropout+XAI)	99.0 ± 0.5	99.0 ± 0.6	100.0 ± 0.0
Without SMOTE (with XAI)	89.0 ± 1.2	89.2 ± 1.1	90.0 ± 1.3
Without Dropout/BatchNorm (with XAI)	90.5 ± 0.8	90.0 ± 0.6	94.0 ± 1.0
Simplified MLP (one hidden layer) (with XAI)	90.0 ± 1.5	90.0 ± 1.4	92.0 ± 1.2
RFEBNet, FCNe, supervised DL [31] (No XAI)	91± 0.3	90± 0.4	93± 0.2
CT-based feature Extraction LASSO selection SVM classifier (No XAI) [33].	86± 0.3	88± 0.1	89± 0.2
Multiple ML classifiers (LR, SVM, RFC, ANN) (No XAI) [36]	70.47± 0.5	70.40± 0.7	80± 0.7

The importance of class imbalance handling in clinical datasets is highlighted by the significant performance decline that results from removing SMOTE. Similarly, the simplified MLP exhibits poorer stability across folds, and eliminating Batch Normalization and Dropout reduces generalization potential. These findings demonstrate that every architectural and preprocessing element significantly improves the proposed framework's overall performance. The proposed algorithm consistently performs better than previous machine-learning techniques [31, 33, 36]. The benefit of jointly learning inspiratory-expiratory imaging and clinical features was demonstrated by the multimodal CNN framework integrating double-phase chest CT and clinical data, which achieved strong diagnostic performance. However, compared to emerging XAI-enabled deep learning approaches, its reliance on dual-phase CT acquisition increases radiation burden, and its lack of explicit explainability limits clinical interpretability [31]. The approach in [33] produced mediocre results (accuracy 86%, macro-F1 88%, ROC-AUC 89%) using a simple CT-based pipeline for feature extraction,

followed by LASSO selection and an SVM classifier. These scores show that although margin-based classifiers can take advantage of carefully chosen radiomic characteristics, they are still not as effective as more expressive ensemble or deep models. Suggesting that while tree-based ensembles may accurately predict clinical characteristics, they are still unable to capture deeper non-linear interactions.

The multi-classifier evaluation in [36] (LR, SVM, RFC, ANN) performed significantly worse (accuracy 70.47%, macro-F1 70.40%, ROC–AUC 80%), underscoring the shortcomings of conventional ML models for multi-class COPD severity prediction.

This study clearly outperforms the conventional machine-learning methods described in the literature, which are restricted to binary or coarse classification, frequently struggle with ambiguous cases, and are unable to capture intricate interactions among clinical factors.

This model minimizes misclassification between clinically adjacent categories, such as mild and moderate COPD, while maintaining high ROC-AUC values across all severity classes, especially in the high-risk severe category, by modeling nonlinear dependencies among symptom scores, smoking history, pulmonary function measures, and demographic features.

Improved generalization and robustness are further enhanced by applying Batch Normalization and Dropout, managing class imbalance with SMOTE, and carrying out thorough ablation and cross-validation tests.

The integration of explainability using SHAP values, which highlight the features impacting each severity prediction, is one of the most important aspects of this work. The explainability study showed that the best signs of COPD severity were spirometry characteristics, particularly FVC, FEV1/FVC ratio, and FEV1 percentage predicted.

This is in line with the GOLD guidelines, which grade airflow limitation using FEV1 and the FEV1/FVC ratio as diagnostic necessities. Significant contributions were also made by longer smoking duration, poorer oxygen saturation, and higher symptom scores, indicating that the model follows clinically proven patterns of COPD progression.

These discoveries improve the model's interpretability and encourage the real-world application of such AI-based systems in healthcare environments where explainability is crucial for clinical decision support. This SHAP analysis offers interpretable insights on feature importance, which can help clinicians interpret and trust model predictions, in contrast to previous ML-based studies [31, 33, 36] that only focus on performance measures.

4.2.5. Implications for Clinical Decision Support

A thorough evaluation framework for COPD severity classification is produced by combining cross-validation, confidence interval reporting, misclassification analysis, and uncertainty estimation.

The suggested solution facilitates human-in-the-loop decision-making by identifying scenarios with high uncertainty and ambiguity. This lowers the risk of relying too much on automated forecasts and enhances clinical interpretability.

4.2.6. Clinical Deployment Considerations

The proposed framework was created with future clinical deployment, even though this study used a de-identified, publicly available dataset. By avoiding the intricacy of imaging pipelines, its emphasis on frequently gathered clinical variables makes it simpler to integrate as a decision-support module within hospital Electronic Health Record (EHR) systems. Practical use requires ethical issues such as avoiding algorithmic bias across demographic groups, patient data protection, and informed consent.

As shown in Figure 10, Gradio was used to create a lightweight web-based user interface that would facilitate clinician contact and show practical application. With the use of this interface, users can enter physiological, symptom-based, and patient demographic data to get class-wise probability estimates and real-time predictions of COPD severity. The UI complements the SHAP-based explainability to improve interpretability and confidence, and it runs solely on CPU resources, making it appropriate for telemedicine platforms and low-resource environments.

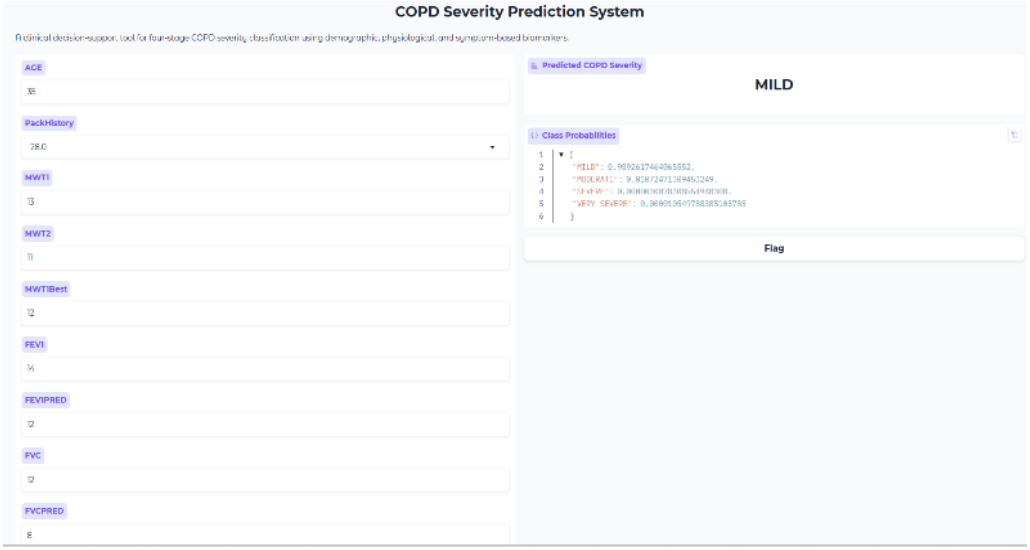
Regulatory compliance and prospective validation will be necessary before real-world deployment. There are practical uses for this research, especially in low-resource or rural healthcare settings where access to professional pulmonologists or imaging equipment may be limited. Because the model just needs patient-reported symptoms, demographic data, and basic spirometry results, it can be used in primary care, telemedicine platforms, and screening camps. It can help non-specialist healthcare providers identify and prioritize COPD patients because of its explainable predictions, robust performance indicators, and SHAP analysis. Despite the great predictive performance and interpretability of the suggested COPD severity categorization system, there are some limitations to the study.

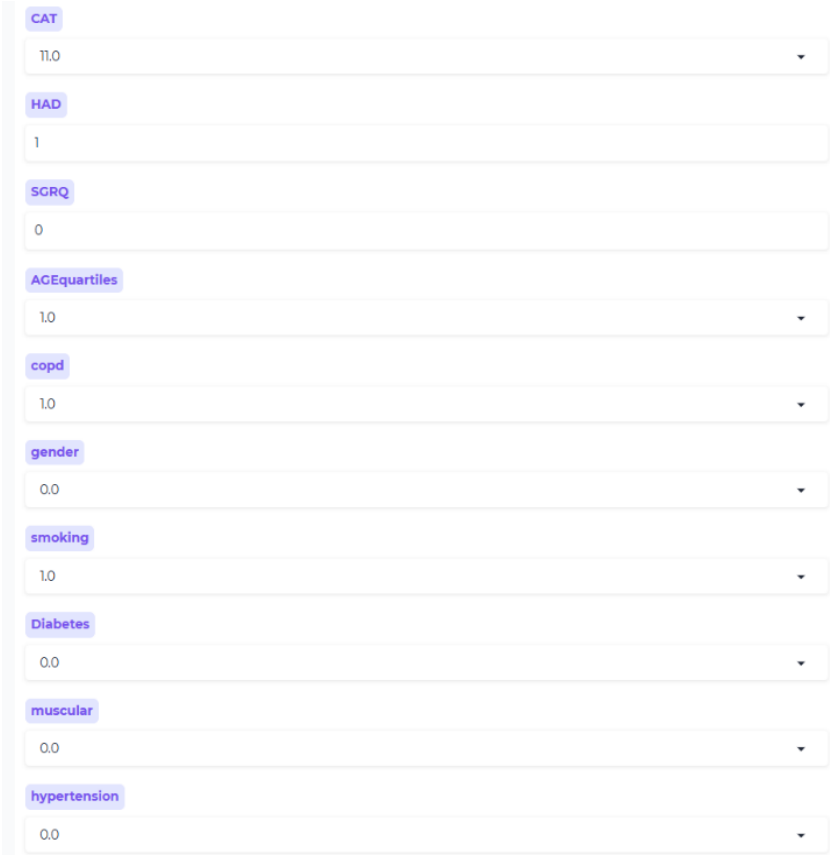
First, the single-center dataset used in this study may limit its applicability to larger populations with distinct clinical protocols or demographics.

Second, rigorous clinical validation across bigger cohorts and multi-center research is still necessary to demonstrate the robustness and therapeutic relevance of these discoveries,

even though SHAP-based explanations offer feature-level interpretability.

Third, imaging or longitudinal data that could improve disease progression modeling and predictive accuracy are now excluded from the model, which only considers tabular clinical data. Fourth, even though stratified cross-validation, SMOTE, and Dropout were used to improve generalization, overfitting could occur due to the dataset's small size. Even though the model performs remarkably well on the test data that is now available, additional external validation on bigger and more varied cohorts will be required to guarantee strong generalizability in actual clinical settings.





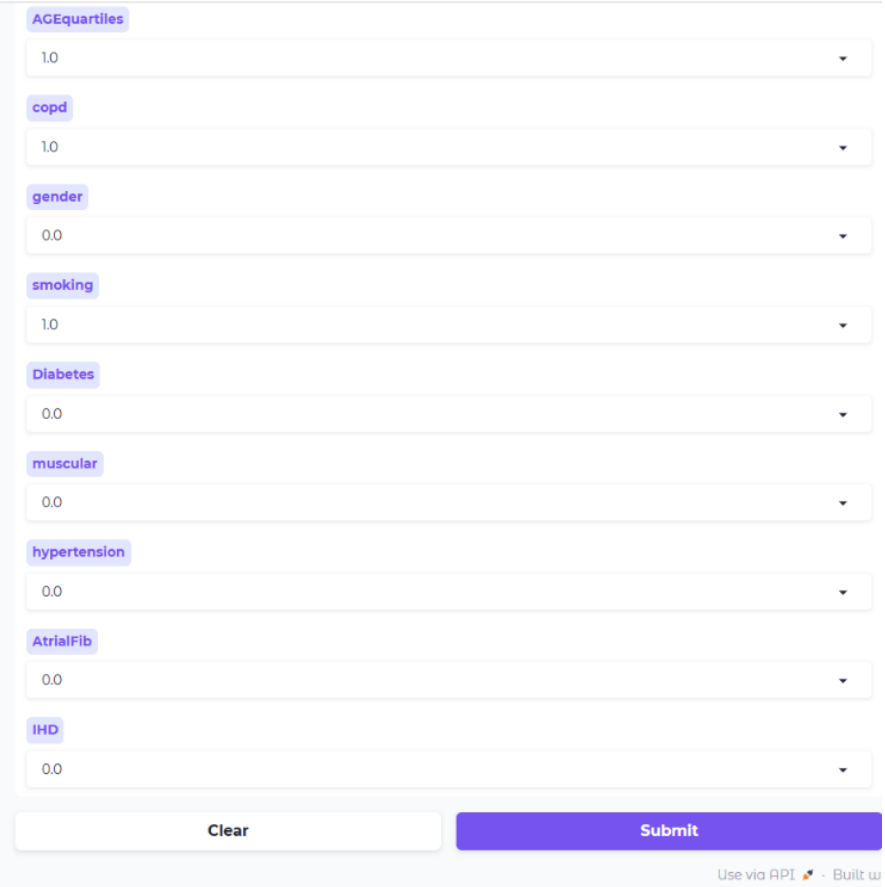


Fig. 10 Screenshot of the COPD severity prediction interface developed with Gradio, showing input fields for clinical features and the predicted severity with probability scores.

5. Conclusion and Future Scope

Using only structured clinical, demographic, and spirometry-derived data, this study presents an explainable deep learning-based model for COPD severity classification. With an excellent mean accuracy of 99.0%, a Macro-F1 score of 98.5%, and an ROC–AUC of 1.00 under five-fold cross-validation, the proposed model showed strong prediction accuracy across mild, moderate, severe, and very severe COPD categories. Ablation and statistical significance analyses further confirm the robustness and reliability of the proposed architecture. These findings demonstrate how clinical data-driven AI models can help control COPD without depending on imaging modalities, making the strategy scalable, affordable, and appropriate for use in healthcare settings with limited resources.

By highlighting physiologically significant characteristics, including FEV1%, FEV1/FVC ratio, FVC, pack history, symptom scores, and demographic risk variables, this work's incorporation of SHAP analysis offers clear insights into the model's decision-making. These results validate the clinical interpretability and reliability of the model in accordance with GOLD recommendations. The use of a single-center dataset, which might not fully represent the diversity of COPD patients, and the emphasis on static severity classification rather than disease progression modeling are two limitations. However, the methodological approach creates a solid basis for use in clinical systems in the future. Future research will focus on developing an early-warning system for exacerbation prediction, external validation using multi-center datasets, and integration with other modalities, such as personal sensor data, imaging features, and cardiac auscultation sounds. Additionally, patient-specific COPD progression patterns may be tracked and predicted over time using an extended deep learning approach. The proposed explainable deep learning framework's clinical dependability, scalability, and translational impact will all be improved by these additions.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding Statement

This research did not receive any specific grant from funding agencies in the public, commercial, or non-profit sectors.

References

[1] Fangfei Wang et al., "Computed Tomography-based Artificial Intelligence in Lung Disease-Chronic Obstructive Pulmonary Disease," *MedComm-Future Medicine*, vol. 3, no. 1, pp. 1-10, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[2] Yaman Akbulut, "Automated Pneumonia based Lung Diseases Classification with Robust Technique based on a Customized Deep Learning Approach," *Diagnostics*, vol. 13, no. 2, pp. 1-17, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[3] Rehan Khan et al., "Auscultation-based Pulmonary Disease Detection through Parallel Transformation and Deep Learning," *Bioengineering*, vol. 11, no. 6, pp. 1-19, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[4] Youssef Abadade et al., "Empowering Healthcare: TinyML for Precise Lung Disease Classification," *Future Internet*, vol. 16, no. 11, pp. 1-14, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[5] Pshtiwan Qader Rashid, and İlker Türker, "Lung Disease Detection using U-Net Feature Extractor Cascaded by Graph Convolutional Network," *Diagnostics*, vol. 14, no. 12, pp. 1-14, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[6] Rima Tri Wahyuningrum et al., "Enhancing Lung Disease Diagnosis with Deep-Learning-based CT Scan Image Segmentation," *Intelligent Systems with Applications*, vol. 27, pp. 1-17, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[7] Md. Nahiduzzaman et al., "Parallel CNN-ELM: A Multiclass Classification of Chest X-Ray Images to Identify Seventeen Lung Diseases Including COVID-19," *Expert Systems with Applications*, vol. 229, pp. 1-19, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[8] Ahmed I. Taloba, and R.T. Matoog, "Detecting Respiratory Diseases using Machine Learning-based Pattern Recognition on Spirometry Data," *Alexandria Engineering Journal*, vol. 113, pp. 44-59, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[9] Goram Mufarah M. Alshmrani et al., "A Deep Learning Architecture for Multi-Class Lung Diseases Classification using Chest X-Ray (CXR) Images," *Alexandria Engineering Journal*, vol. 64, pp. 923-935, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[10] World Health Organization, World Health Statistics 2024: Monitoring Health for the SDGs, Sustainable Development Goals, Global Report, 2024. [Online]. Available: https://www.who.int/publications/i/item/9789240094703

[11] Zhuoneng Zhang et al., "Detection of Chronic Obstructive Pulmonary Disease with Deep Learning using Inspiratory and Expiratory Chest Computed Tomography and Clinical Information," *Journal of Thoracic Disease*, vol. 16, no. 9, pp. 6101-6111, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[12] Johan N. Siebert et al., "Deep Learning Diagnostic and Severity-Stratification for Interstitial Lung Diseases and Chronic Obstructive Pulmonary Disease in Digital Lung Auscultations and Ultrasonography: Clinical Protocol for an Observational Case-Control Study," *BMC Pulmonary Medicine*, vol. 23, no. 1, pp. 1-6, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[13] Alvar Agusti et al., "Global Initiative for Chronic Obstructive Lung Disease 2023 Report: GOLD Executive Summary," *American Journal of Respiratory and Critical Care Medicine*, vol. 207, no. 7, pp. 819-837, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[14] Al-Anazi Saad et al., "Artificial Intelligence in Respiratory Care: Current Scenario and Future Perspective," *Annals of Thoracic Medicine*, vol. 19, no. 2, pp. 117-130, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[15] Xiuying Mou et al., "Detection and Airflow Obstruction Severity Grading of Chronic Obstructive Pulmonary Disease based on Capnography," *Nanotechnology Precision Engineering*, vol. 8, no. 3, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[16] Nicole M Robertson, Connor S Centner, and Trishul Siddharthan, "Integrating Artificial Intelligence in the Diagnosis of COPD Globally: A Way Forward," *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation*, vol. 11, no. 1, pp. 114-120, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[17] Xueting Shen, and Huanbing Liu, "Using Machine Learning for Early Detection of Chronic Obstructive Pulmonary Disease: A Narrative Review," *Respiratory Research*, vol. 25, no. 1, pp. 1-9, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[18] Laurie D Snyder et al., "Predictive Machine Learning Algorithm for COPD Exacerbations using a Digital Inhaler with Integrated Sensors," *BMJ Open Respiratory Research*, vol. 12, no. 1, pp. 1-12, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[19] Luke A Smith et al., "Machine Learning and Deep Learning Predictive Models for Long-Term Prognosis in Patients with Chronic Obstructive Pulmonary Disease: A Systematic Review and Meta-Analysis," *The Lancet Digital Health*, vol. 5, no. 12, pp. E872-E881, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[20] Siyang Zeng, Mehrdad Arjomandi, and Gang Luo, "Automatically Explaining Machine Learning Predictions on Severe Chronic Obstructive Pulmonary Disease Exacerbations: Retrospective Cohort Study," *JMIR Medical Informatics*, vol. 10, no. 2, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[21] XiaoLing Zou et al., "Screening and Staging of Chronic Obstructive Pulmonary Disease with Deep Learning based on Chest X-Ray Images and Clinical Parameters," *BMC Pulmonary Medicine*, vol. 24, no. 1, pp. 1-10, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[22] Ali Hussain et al., "Forecast the Exacerbation in Patients of Chronic Obstructive Pulmonary Disease with Clinical Indicators using Machine Learning Techniques," *Diagnostics*, vol. 11, no. 5, pp. 1-17, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[23] Zian Liu et al., "Construction of a Severity Prediction Model for Hospitalized Patients with Acute Exacerbation of Chronic Obstructive Pulmonary Disease based on Machine Learning," *European Journal of Medical Research*, vol. 31, no. 1, pp. 1-15, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[24] Erkang Yi et al., "An Integrated Machine Learning Model of Transcriptomic Genes in Multi-Center Chronic Obstructive Pulmonary Disease Reveals the Causal Role of TIMP4 in Airway Epithelial Cell," *Respiratory Research*, vol. 26, no. 1, pp. 1-22, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[25] Uppuluri Ruchitha Venkata Sai Meenakshi, and Vikas Jindal, "A Hybrid Feature Selection Model for Predicting Chronic Obstructive Pulmonary Disease," *Third International Conference on Inventive Research in Computing Applications (ICIRCA)*, Coimbatore, India, pp. 589-596, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[26] Shuhao Mei et al., "Deep Learning for Detecting and Early Predicting Chronic Obstructive Pulmonary Disease from Spirogram Time Series," *npj Systems Biology and Applications*, vol. 11, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[27] Seung Yeob Ryu et al., "Machine Learning-based Prediction of in-Hospital Mortality in Patients with Chronic Respiratory Disease Exacerbations," *Digital Health*, vol. 11, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[28] Genevera I. Allen, Luqin Gan, and Lili Zheng, "Interpretable Machine Learning for Discovery: Statistical Challenges and Opportunities," *Annual Review of Statistics and its Application*, vol. 11, no. 1, pp. 77-101, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[29] Scott M Lundberg, and Su-In Lee, "A Unified Approach to Interpreting Model Predictions," *Advances in Neural Information Processing Systems*, vol. 30, 2017. [Google Scholar] [Publisher Link]

[30] Israa Mohamed, "Prediction of Chronic Obstructive Pulmonary Disease Stages Using Machine Learning Algorithms," *International Journal of Decision Support System Technology*, vol. 14, no. 1, pp. 1–13, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[31] Zhuoneng Zhang et al., "Detection of Chronic Obstructive Pulmonary Disease with Deep Learning using Inspiratory and Expiratory Chest CT and Clinical Information," *Journal of Thoracic Disease*, vol. 16, no. 9, pp. 6101-6111, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[32] Nian Cai et al., "Deep Learning Assisted Diagnosis of Chronic Obstructive Pulmonar Disease based on a Local-to-Global Framework," *Electronics*, vol. 13, no. 22, pp. 1-13, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[33] He Sui et al., "Diagnosis and Severity Assessment of COPD based on Machine Learning of Chest CT Images," *International Journal of Chronic Obstructive Pulmonary Disease*, vol. 20, pp. 2853-2867, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[34] Chew-Teng Kor et al., "Explainable Machine Learning Model for Predicting First-Time Acute Exacerbation in Patients with Chronic Obstructive Pulmonary Disease," *Journal of Personalized Medicine*, vol. 12, no. 2, pp. 1-19, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[35] Arnold Kamis et al., "Obtaining the Most Accurate, Explainable Model for Predicting Chronic Obstructive Pulmonary Disease: Triangulation of Multiple Linear Regression and Machine Learning Methods," *JMIR AI*, vol. 3, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[36] Walaa H. Elashmawi et al., "Machine Learning for Enhanced COPD Diagnosis: A Comparative Analysis of Classification Algorithms," *Diagnostics*, vol. 14, no. 24, pp. 1-26, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[37] Xiuying Mou et al., "A Novel Approach for the Detection and Severity Grading of Chronic Obstructive Pulmonary Disease based on Transformed Volumetric Capnography," *Bioengineering*, vol. 11, no. 6, pp. 1-15, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[38] Prakhar Rathi, COPD Patients Dataset, Data of Physical Characteristics of Patients with COPD, 2021. [Online]. Available: https://www.kaggle.com/datasets/prakharrathi25/copd-student-dataset

[39] GOLD Report 2024: Global Initiative for Chronic Obstructive Lung Disease, Global Strategy for the Diagnosis, Management, and Prevention of COPD, 2024. [Online]. Available: https://goldcopd.org/2024-gold-report/

[40] Deepshikha Bhati, Fnu Neha, and Md Amiruzzaman, "A Survey on Explainable Artificial Intelligence (XAI) Techniques for Visualizing Deep Learning Models in Medical Imaging," *Journal of Imaging*, vol. 10, no. 10, pp. 1-26, 2024. [CrossRef] [Google Scholar] [Publisher Link]