

Uncovering Clinical Subtypes of Interstitial Lung Disease through latent class analysis, an unsupervised Machine Learning algorithm: Reg No: 418

Ankita Dey, Vikram Vohra, Priyanshi Joshi
LRS Institute of Tuberculosis and Respiratory Diseases

CORRESPONDING AUTHOR

Dr Ankita Dey, LRS Institute of Tuberculosis and Respiratory Diseases
Email: ankitadey14@gmail.com

CITATION

Dey Ankita, Vohra Vikram, JoshiPriyanshi. Uncovering Clinical Subtypes of Interstitial Lung Disease through latent class analysis, an unsupervised Machine Learning algorithm: Reg No: 418. Journal of the Epidemiology Foundation of India. 2025;3(2Suppl):25.

DOI: <https://doi.org/10.56450/JEFI.2025.v3i2Suppl.025>

ARTICLE CYCLE

Received: 10/11/2025; Accepted: 20/12/2026; Published: 31/12/2025

This work is licensed under a Creative Commons Attribution 4.0 International License.

©The Author(s). 2026 Open Access

ABSTRACT

Introduction: Interstitial lung diseases (ILDs) represent a heterogeneous group of pulmonary disorders characterized by varying clinical presentations, prognoses, and responses to treatment. Traditional classification methods often fail to capture the complexity and diversity within the ILD patient population. To address this challenge, data-driven approaches such as unsupervised machine learning offer promising avenues for subgroup identification. **Objective:** This study explores the application of an unsupervised machine learning technique, Latent Class Analysis (LCA), to identify clinically meaningful subgroups of ILD patients based on patterns of signs and symptoms. **Methods:** Clinical data from a cohort of ILD patients were analyzed, focusing on key symptoms (e.g., dyspnea, cough, fatigue) and clinical indicators. LCA was applied to uncover latent subgroups without predefined labels. Model selection was guided by statistical fit indices (e.g., BIC, AIC) and clinical interpretability. **Results:** LCA revealed distinct subgroups of ILD patients, each characterized by unique symptom profiles and clinical features. These subgroups demonstrated significant differences in disease severity, functional status, and potential prognostic outcomes, suggesting that symptom-based clustering can provide meaningful stratification beyond conventional diagnostic categories. **Conclusion:** Latent Class Analysis effectively identified subpopulations within the ILD patient group based on symptomatology. These findings highlight the potential of unsupervised machine learning to enhance clinical phenotyping, guide personalized management strategies, and inform future research on ILD heterogeneity.

KEYWORDS

Interstitial lung disease, Latent class analysis, Machine learning, Symptom clusters, Patient stratification, Unsupervised learning