Supplementary Table S1. Interobserver agreement for predominant HRCT pattern and individual imaging features

HRCT variable	Observed agreement (%)	Cohen's κ	95% CI
Predominant HRCT pattern	100.0	1.000	0.888–1.000
Craniocaudal distribution	98.4	0.952	0.885–1.000
Axial distribution	97.6	0.963	0.921–1.000
Symmetry	100.0	1.000	0.823–1.000
Traction bronchiectasis	98.4	0.967	0.922–1.000
Traction bronchiolectasis	98.4	0.891	0.743–1.000
Honeycombing	99.2	0.975	0.927–1.000
Architectural distortion	99.2	0.983	0.950–1.000
Ground-glass opacity	99.2	0.952	0.859–1.000
Consolidation	98.4	0.954	0.890–1.000
Cysts	99.2	0.983	0.949–1.000
Nodules	100.0	1.000	0.823–1.000
Centrilobular ground-glass nodules	96.7	0.924	0.851–0.997
Mosaic attenuation	100.0	1.000	0.823–1.000
Three-density sign	100.0	1.000	0.823–1.000

Abbreviations: HRCT, high-resolution computed tomography; CI, confidence interval.

Note: Interobserver agreement was calculated using Cohen's kappa coefficient. For five variables with perfect observed agreement (predominant HRCT pattern, symmetry, nodules, mosaic attenuation, and three-density sign), the 95% confidence intervals were obtained using an approximate fallback procedure based on the null-hypothesis standard error because the primary procedure could not provide a finite standard error. These confidence intervals should therefore be interpreted with caution.

Supplementary Table S2. Comparison of patients with and without available pulmonary function test data.

Characteristic	PFT available (n=106)	PFT unavailable (n=17)	p-value
Age at HRCT, years, median (Q1–Q3)	60 (49–66)	65 (57–73)	0.055

Sex, n (%)			1.000
Female	75 (70.8%)	12 (70.6%)	
Male	31 (29.2%)	5 (29.4%)	
Underlying CTD, n (%)			0.836
Systemic sclerosis	47 (44.3%)	8 (47.1%)	
Rheumatoid arthritis	30 (28.3%)	6 (35.3%)	
Idiopathic inflammatory myopathies	10 (9.4%)	0 (0%)	
Sjögren's syndrome	7 (6.6%)	1 (5.9%)	
Mixed connective tissue disease	3 (2.8%)	1 (5.9%)	
Systemic lupus erythematosus	3 (2.8%)	0 (0%)	
Undifferentiated CTD	6 (5.7%)	1 (5.9%)	
Predominant HRCT pattern, n (%)			0.896
UIP	23 (21.7%)	4 (23.5%)	
NSIP	60 (56.6%)	9 (52.9%)	
OP	1 (0.9%)	0 (0%)	
LIP	1 (0.9%)	0 (0%)	
BIP	12 (11.3%)	3 (17.6%)	
PPFE	0 (0%)	0 (0%)	
Mixed	0 (0%)	0 (0%)	
Unclassifiable	9 (8.5%)	1 (5.9%)	
Fibrosis extent, median (Q1–Q3)	2 (2–3)	2 (1–3)	0.835
Cluster membership, n (%)			0.667
Cluster 1	47 (44.3%)	6 (35.3%)	
Cluster 2	22 (20.8%)	5 (29.4%)	
Cluster 3	37 (34.9%)	6 (35.3%)	

Values are presented as median (Q1–Q3) or number (percentage), as appropriate. Percentages are calculated within each PFT availability group. Age and fibrosis extent were compared using the Mann–Whitney U test. Categorical variables were compared using Fisher's exact test.

Abbreviations: BIP, bronchiolocentric interstitial pneumonia; CTD, connective tissue disease; HRCT, high-resolution computed tomography; IIM, idiopathic inflammatory myopathies; LIP, lymphoid interstitial pneumonia; NSIP, nonspecific interstitial pneumonia; OP, organizing pneumonia; PFT, pulmonary function tests; PPFE, pleuroparenchymal fibroelastosis; UIP, usual interstitial pneumonia.

Supplementary File S1. R code for the reproducibility and bootstrap stability assessment of the MCA–HCPC clustering procedure.

This supplementary file provides the R code used to reproduce the three-cluster solution obtained through multiple correspondence analysis (MCA) and hierarchical clustering on principal components (HCPC), and to assess internal cluster stability using bootstrap resampling. The script represents an edited extraction of the original analysis code, retaining the relevant analytical procedures, software package versions, and statistical parameters. The code is provided in Word (.docx) format to facilitate accessibility and review. No patient-level data or individual cluster assignments are included. The original study dataset is not publicly available due to patient privacy and ethical restrictions, as stated in the manuscript.

```
# Supplementary R code: MCA-HCPC reproducibility and bootstrap stability
# Manuscript: HRCT Imaging Phenotypes and Bronchiolocentric Involvement in
# Connective Tissue Disease-Associated Interstitial Lung Disease
#
# This is an edited extraction of the original analysis script. It retains
# the MCA-HCPC implementation and clusterboot settings used for the
# reported three-cluster solution, without unrelated exploratory analyses.
#
# INPUT FILES (provided by the researcher; NOT included in this script):
#   1. HRCT_data.xlsx: one row per patient; columns Study_ID and the 14
#      categorical HRCT variables below, coded exactly as in the analysis.
#   2. original_cluster_assignments.csv: Study_ID,Cluster_original.
#      This file is optional for the bootstrap itself, but needed to verify
#      concordance with the original patient assignments (ARI).
#
# IMPORTANT: Study_ID serves only to align rows and is NOT included in MCA.
# Original software: MEDA 1.6.0 / FactoMineR 2.11 in jamovi.
# R reproduction environment:
# R version 4.6.2
# FactoMineR version 2.18
# mclust version 6.1.3
# fpc version 2.2.15
# Three clusters are FIXED here (nb.clust = 3), mirroring the published
```

```
# partition. This script does not independently validate automatic choice
# of the number of clusters in other software versions.

required_packages <- c('readxl', 'FactoMineR', 'fpc', 'mclust')
missing_packages <- required_packages[!vapply(required_packages, requireNamespace,
                                  logical(1), quietly = TRUE)]
if (length(missing_packages)) {
  stop('Install required R packages first: ', paste(missing_packages, collapse = ', '))
}

# Select the original HRCT dataset interactively (not included).
hrct_file <- file.choose()
# Optional: set the path to a separate file of original cluster assignments
# to enable ARI comparison. The script also runs without this file.
cluster_file <- 'original_cluster_assignments.csv'

raw_data <- as.data.frame(readxl::read_excel(hrct_file))
hrct_vars <- c(
  'Craniocaudal_Distribution', 'Axial_Distribution', 'Symmetry',
  'Traction_Bronchiectasis', 'Traction_Bronchiolectasis',
  'Honeycombing', 'Architectural_Distortion', 'Ground_Glass_Opacity',
  'Consolidation', 'Cysts', 'Nodules', 'Centrilobular_GGN',
  'Mosaic_Attenuation', 'Three_Density_Sign'
)

stopifnot('Study_ID' %in% names(raw_data),
          all(hrct_vars %in% names(raw_data)),
          !anyNA(raw_data$Study_ID),
          !anyDuplicated(raw_data$Study_ID))

# Factor conversion must preserve the original category coding.
date_MCA <- as.data.frame(raw_data[, hrct_vars, drop = FALSE])
date_MCA[] <- lapply(date_MCA, factor)
if (anyNA(date_MCA)) stop('Missing HRCT values: check preprocessing before MCA.')

# 1. Reproduce the original five-dimensional MCA-HCPC partition.
set.seed(123)
MCA_original_reprodus <- FactoMineR::MCA(
  date_MCA,
  ncp = 5,
  level.ventil = 0.05,
  graph = FALSE
)
HCPC_trei_test <- FactoMineR::HCPC(
  MCA_original_reprodus,
  nb.clust = 3,
  metric = 'euclidean',
  method = 'ward',
```

```
  consol = TRUE,
  graph = FALSE
)
reproduced_clusters <- as.integer(as.character(HCPC_trei_test$data.clust$clust))
cat('Reproduced cluster sizes:\n')
print(table(reproduced_clusters))

# Compare with original patient labels, if provided.
if (file.exists(cluster_file)) {
  original <- read.csv(cluster_file, stringsAsFactors = FALSE)
  stopifnot(all(c('Study_ID', 'Cluster_original') %in% names(original)),
            !anyDuplicated(original$Study_ID))
  idx <- match(raw_data$Study_ID, original$Study_ID)
  if (anyNA(idx)) stop('Some Study_ID values are missing from original labels.')
  original_clusters <- original$Cluster_original[idx]
  cat('\nOriginal versus reproduced cluster assignments:\n')
  print(table(Original = original_clusters, Reproduced = reproduced_clusters))
  cat('\nAdjusted Rand Index (original vs reproduced): ',
      mclust::adjustedRandIndex(original_clusters, reproduced_clusters), '\n', sep = '')
} else {
  message('Original cluster file not found: skipping original-versus-reproduced ARI.')
}

# 2. Bootstrap: retain the ORIGINAL user-supplied custom callback and settings.
# The callback reruns MCA and HCPC in every bootstrap resample.
# Resampled categorical columns with a single observed level are removed.
HCPC_original_CBI <- function(data, k, ...) {
  x <- droplevels(as.data.frame(data))
  x <- x[, sapply(x, nlevels) > 1, drop = FALSE]

  mca <- FactoMineR::MCA(
    x,
    ncp = 5,
    level.ventil = 0.05,
    graph = FALSE
  )

  hcpc <- FactoMineR::HCPC(
    mca,
    nb.clust = k,
    metric = 'euclidean',
    method = 'ward',
    consol = TRUE,
    graph = FALSE
  )

  grupuri <- as.integer(hcpc$data.clust$clust)
  list(
```

```
    result = hcpc,
    nc = k,
    clusterlist = lapply(seq_len(k), function(j) grupuri == j),
    partition = grupuri,
    clustermethod = 'MCA + HCPC'
  )
}

set.seed(123)
validare_originala <- fpc::clusterboot(
  date_MCA,
  B = 1000,
  bootmethod = 'boot',
  clustermethod = HCPC_original_CBI,
  k = 3,
  datatomatrix = FALSE,
  multipleboot = TRUE,
  bscompare = TRUE,
  dissolution = 0.50,
  recover = 0.75,
  seed = 123,
  count = FALSE
)

cat('\nBootstrap stability report:\n')
print(validare_originala)
cat('\nClusterwise mean Jaccard similarities:\n')
print(validare_originala$bootmean)
cat('\nR session and package versions:\n')
print(sessionInfo())

# IMPORTANT: Cluster numbers are arbitrary labels. For comparisons with
# published C1/C2/C3 values, verify the correspondence of bootstrap
# reference clusters with the original patient assignment file.
```