

Article

Not peer-reviewed version

Utility of Clustering in Mortality Risk Stratification in Pulmonary Hypertension

[Pasquale Tondo](#) , [Lucia Tricarico](#) , [Giuseppe Galgano](#) , Maria Pia Varlese , Daphne Aruanno , [Crescenzo Gallo](#) ^{*} , Giulia Scioscia , [Natale Brunetti](#) , [Michele Correale](#) , [Donato Lacedonia](#)

Posted Date: 18 March 2025

doi: 10.20944/preprints202503.1369.v1

Keywords: Pulmonary Hypertension; Clustering; Machine Learning



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Utility of Clustering in Mortality Risk Stratification in Pulmonary Hypertension

Pasquale Tondo, Lucia Tricarico, Giuseppe Galgano, Maria Pia Varlese, Daphne Aruanno, Crescenzo Gallo *, Giulia Scioscia, Natale Brunetti, Michele Correale and Donato Lacedonia

Affiliation

* Correspondence: crescenzo.gallo@unifg.it

Abstract: **INTRODUCTION:** Pulmonary hypertension (PH) is a condition characterized by increased pressure in the pulmonary arteries with poor prognosis and, so, an optimal management is necessary. **OBJECTIVES:** The study aim was to search for PH phenotypes and develop a predictive model of five-year mortality using machine learning (ML) algorithms. **METHODS:** This multicenter study was conducted on 122 PH patients. Clinical and demographic data were collected and, then, used to identify phenotypes through clustering. Subsequently, predictive model was performed by different ML algorithms. **RESULTS:** Three PH clusters were identified: Cluster 1 (mean age $68,57 \pm 10,54$) includes 57% females, 69% from non-respiratory PH groups, and better cardiac (NYHA class $2,61 \pm 0,84$) and respiratory function (FEV1% $78,78 \pm 21,54$); Cluster 2 includes 50% females, mean age of 71.36 ± 8.32 years, 44% from PH group 3, worse respiratory function (FEV1% $68,12 \pm 10,20$), intermediate cardiac function (NYHA class $3,18 \pm 0,49$) and significantly higher mortality (75%); Cluster 3 represents the youngest cluster (mean age $61,11 \pm 13,50$) with 65% males, 81% from non-respiratory PH groups, intermediate respiratory function (FEV1% $70,51 \pm 17,91$) and worse cardiac performance (NYHA class $3,22 \pm 0,58$). After testing ML models, logistic regression showed the best predictive performance (AUC=0.835 and accuracy=0.744.) and identified three mortality-risk factors: age, NYHA class, number of medications taken. **CONCLUSIONS:** The results suggest that the integration of ML into clinical practice can improve risk stratification to optimize treatment strategies and improve the outcome of PH patients.

Keywords: pulmonary hypertension; clustering; machine learning

1. Introduction

Pulmonary hypertension (PH) is a rare heterogeneous disease characterized by elevated blood pressure in the pulmonary circulation, caused by both pulmonary vascular remodeling and inflammation or increased downstream pressures [1]. This condition ultimately results in right heart failure and significantly contributes to morbidity and mortality among affected patients. The complexity of PH arises from its diverse etiologies, including idiopathic origins, connective tissue diseases, congenital heart defects, left heart disease, lung disease and chronic thromboembolic events, which are categorized into five distinct clinical groups by the World Health Organization (WHO)1: pulmonary arterial hypertension (PAH) (group 1), pulmonary hypertension due to left heart disease (PH-LHD) (group 2), pulmonary hypertension due to chronic lung disease (PH-CLD) and/or hypoxia (group 3), pulmonary hypertension associated with pulmonary artery obstructions (group 4), pulmonary hypertension with unclear and/or multifactorial mechanisms (group 5).

The hemodynamic definition of PH is the elevation of mean pulmonary arterial pressure (mPAP) >20 mmHg, assessed by right heart catheterization (RHC). Pulmonary arterial wedge pressure (PAWP) and pulmonary vascular resistance (PVR) distinguish pre-capillary PH (PAWP ≤ 15 mmHg, PVR >2 Wood Units (WU)), isolated post-capillary PH (PAWP >15 mmHg, PVR ≤ 2 WU) and combined post- and pre-capillary PH (PAWP >15 mmHg, PVR >2 WU)1, [2].

PH affects all age groups, with an estimated prevalence of 1% of the world's population. The prevalence is higher

in patients aged > 65 years due to increased cardiac and pulmonary etiologies, with lung disease and chronic

obstructive pulmonary disease (COPD) being a common cause of PH1.

Patients usually present with non-specific symptoms, such as dyspnea on exertion or at rest, fatigue and rapid exhaustion, bendopnea, angina, syncope and/or clinical signs of right heart failure [3].

Despite advancements in therapeutic options over the past decades, PH remains challenging to diagnose and manage due to its heterogeneous nature and the variability in disease progression among patients. Accurate and early diagnosis, coupled with effective management strategies, are crucial for improving patient outcomes and quality of life.

However, underdiagnosis PH is a significant issue in clinical settings. Various factors impede the timely detection of this condition: diagnostic delay, due to the lack of specific symptoms and/or of awareness among general practitioners that can result in delayed referrals to specialized PH centers, limited access to diagnostic procedures, variability in clinical practices and interpretation of results.

Currently, there are only two validated screening algorithms for pulmonary hypertension (PH): one for PAH secondary to scleroderma [4] and another for chronic thromboembolic pulmonary hypertension (CTEPH) following pulmonary embolism [5].

The prognosis of pulmonary hypertension is variable and depends on a mixture of hemodynamic, clinical, etiopathologic parameters and serum biomarkers. Early diagnosis, accurate risk stratification, and appropriate treatment are essential to improve outcomes. The management of PH requires a multidisciplinary approach in specialized centers, with continuous assessment of treatment response and adjustment of therapy based on the risk profile.

Artificial intelligence (AI), particularly machine learning (ML), has shown promise in predicting cardiac conditions, including PAH [6,7]. Given that PAH patients tend to have high healthcare resource utilization before diagnosis [8], ML could potentially use this data for early detection [9].

This study aims to identify homogeneous patient clusters within a PH population and develop a predictive model for 5-year mortality, using ML techniques, thereby enhancing clinical decision-making and patient management.

2. Methods

2.1. Study Population

A prospective study was conducted on 142 patients evaluated for suspected PH by expert cardiologists and pneumologists from the University-Hospital Polyclinic of Foggia and the Miulli Hospital of Acquaviva delle Fonti between 2009 and 2018. PH diagnosis was confirmed through RHC, and patients were followed for five years. During baseline evaluation, we collected several data including demographics (gender, age), clinical information (date of first diagnosis, PH group, NYHA classification, and number of medications), results from spirometry, DL_{CO}, 6 Minute Walking Test (6MWT), RHC and, at 5-year from diagnosis, survival rate was assessed.

2.2. Cardio-Pulmonary Tests

Spirometry

Spirometry was performed using standardized equipment and technique, as defined by the American Thoracic Society/European Respiratory Society (ATS/ERS) task force [10]. All applicants performed three forced vital capacity (FVC) maneuvers, and the best of the 3 measurements was recorded. We collected the following spirometry data: FVC, forced expiratory volume in one second (FEV1), peak expiratory flow (PEF), Tiffeneau index (FEV1/FVC ratio), and forced expiratory flow

during the middle half of the FVC maneuver (FEF25-75%). All spirometry values were corrected for height, weight, age, and sex.

DL_{CO}

Diffusion within the lungs is an electrochemical process that occurs between the gas and liquid phases, driven by the gradient of partial pressures of the gases involved. The diffusion capacity quantifies the volume of any gas (mL) that passes through the alveolocapillary membrane per unit time (in minutes) under a specified pressure difference (1 mmHg). This capacity is expressed in units of mL/min/mmHg. The measurement is conducted with hemoglobin correction in accordance with the standards established by the European Respiratory Society and the American Thoracic Society (ERS/ATS) [11].

Six Minute Walking Test

Six minutes walking test (6MWT) was performed according to ATS 2002 guidelines for all patients as a marker of exercise tolerance with assessment of distance a patient can walk in six minutes and desaturation difference between the baseline SpO₂ and post exercise SpO₂ [12].

Right Heart Catheterization

Hemodynamic assessment was conducted using a Swan-Ganz catheter (CCOmbo V, Edwards Lifesciences). Measurements of systolic, diastolic, and mPAP, right atrial pressure (RAP), and pulmonary capillary wedge pressure were taken at the end of a quiet respiratory cycle. Oxygen saturations were recorded in the superior vena cava, inferior vena cava, pulmonary artery, and femoral artery, with pulmonary vein saturation assumed at 98%. Pulmonary and systemic flows were calculated using the Fick principle. Pulmonary and systemic vascular resistance indexes were determined using standard formulas. A pulmonary arterial wedge pressure greater than 15 mm Hg excluded the diagnosis of precapillary PAH.

2.3. Data Analysis

The statistical analysis was conducted using IBM SPSS Statistics version 26 (IBM Software, CA, USA) [13]. The distribution of the sample was analyzed by Shapiro-Wilk test. Continuous variables were presented as mean and standard deviation (SD), while categorical variables were expressed as percentages. The data included in the analysis were used to identify homogeneous subgroups of patients (phenotypes) through clustering. Differences between clusters were assessed using one-way ANOVA, with post-hoc Tukey analysis validating the results. Mortality among the phenotypes was represented by Kaplan-Meier curves and analyzed using the log-rank test. A p-value of less than 0.05 was considered statistically significant.

Unsupervised and supervised analysis

ML analyses (clustering and predictive methods) were performed using the Orange software [14].

Patients with confirmed diagnosis of PH were considered for analysis, and all collected variables were utilized. Through hierarchical clustering three homogeneous subgroups (clusters) were identified by means of Ward's linkage method. The number of clusters was confirmed by the Silhouette plot index, calculated using Manhattan distance.

Subsequently, the Fast Correlation Based Filter (FCBF) ranking method was employed for searching the main subgroups features outcome-related. This ranking method identified three features that best determined the outcome of this study: the number of medications, age, and NYHA class.

After features selection, various ML algorithms were used to develop a model for predicting mortality in PH: Logistic Regression, Support Vector Machine (SVM), Random Forest, and Neural Network.

The results of the models obtained from the four selected algorithms were validated using 10-Fold Cross Validation.

The effectiveness of the models was evaluated using Area under the Receiver Operating Characteristic (ROC) curve (AUC) and Classification Accuracy (CA). Based on these metrics, Logistic

Regression was found to be the most effective method (AUC = 0.835 and CA = 0.744) for the target variable (**Figure 1**), represented by 5-year survival from diagnosis.

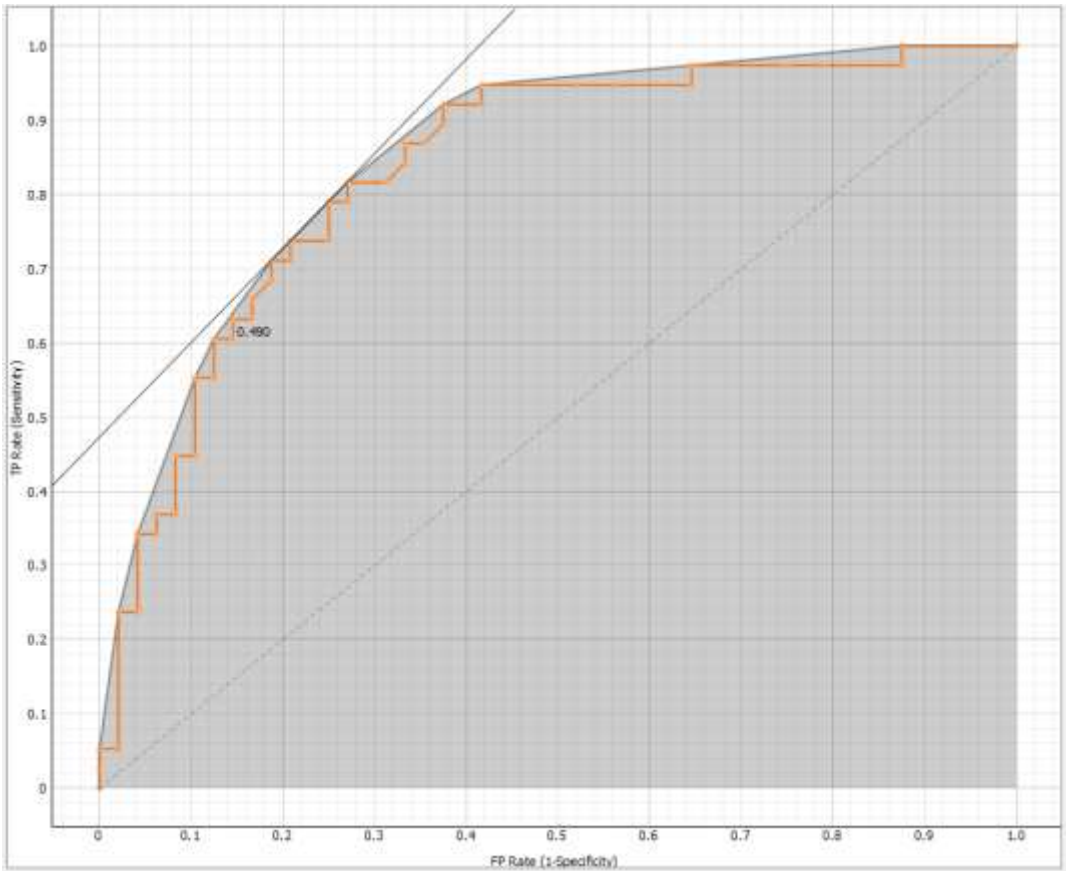


Figure 1. ROC Curve for logistic regression.

3. Results

Out of 142 patients, 122 with confirmed PH were included in the final analysis. Population characteristics are listed in **Table 1**.

Table 1. Population’s characteristics.

Clinical Parameters	N=122
Age (years)	67,13±11,64
Male (%)	52%
PH Group 1,2,4,5 (%)	69%
PH Group 3 (%)	31%
FVC (%)	77,56±19,10
FEV ₁ (%)	73,12±18,23
FEV ₁ /FVC (%)	76,40±13,20
DL _{CO} (%)	48,20±18,92
DL _{CO} /V _A (%)	74,18±21,95
FVC/DL _{CO} (%)	1,84±0,67
sPAP (mmHg)	68,72±21,95
mPAP (mmHg)	39,46±12,46
PCWP (mmHg)	14,60±6,40
CI (Lmin ⁻¹ mt ⁻²)	2,69±1,09
RAP (mmHg)	11,55±6,26
PVR (Wu)	5,83±3,86
TPG (mmHg)	24,83±13,46

6MWT_(mt)	276,21±119,31
WHO functional class (n)	2,96±0,73
Number of drugs (n)	1,25±0,98
Exitus (%)	47%
Months of survival (n)	43,03±22,25

The study involved patients with a mean age of 67.13 ± 11.64 years, of which 52% were male and 48% were female. We divided the patients into two main groups based on the 2022 ESC/ERS guidelines: 31% to "Group 3" (PH associated with lung diseases and/or hypoxia) and 69% belonged to the "Other Group" (patients from groups 1, 2, 4, and 5 of the clinical classification of PH).

The average pulmonary function values showed an FVC of 77.56 ± 19.10 % of the predicted value and an FEV₁ of 73.12 ± 18.23 % of the predicted value, with an FEV₁/FVC ratio of 76.40 ± 13.20 . The DL_{CO} was 48.20 ± 18.92 % of the predicted value, while the DL_{CO}/VA ratio was 74.18 ± 21.95 . Hemodynamic parameters indicated a sPAP of 68.72 ± 21.95 mmHg and a mPAP of 39.46 ± 12.46 mmHg. Exercise capacity, evaluated with the 6MWT, showed a mean distance covered of 276.21 ± 119.31 meters, with most patients classified in WHO functional class II-III. Finally, the average number of medications taken was 1.25 ± 0.98 , and the reported mortality rate was 47%, with a mean survival of 43.03 ± 22.25 months. By dividing the patients into three clusters, it was possible to make comparisons based on different clinical and demographic characteristics (**Table 2**).

Table 2. Comparison between clusters.

	Cluster 1 N=49	Cluster 2 N=36	Cluster 3 N=37	P
Age (years)	68,57±10,54 ^b	71,36±8,32 ^c	61,11±13,50 ^{b, c}	<0,001
Male (%)	43%	50%	65%	0,128
PH Group 1,2,4,5 (%)	69%	56%	81%	0,063
PH Group 3 (%)	31%	44%	19%	0,063
FVC (%)	83,34±19,40 ^a	68,01±12,66 ^{a, c}	79,18±20,71 ^c	0,001
FEV ₁ (%)	78,78±21,54 ^a	68,12±10,20 ^a	70,51±17,91	0,015
FEV ₁ /FVC (%)	76,15±13,29	82,33±9,98 ^c	70,96±13,70 ^c	0,001
DL _{CO} (%)	58,82±20,91 ^{a, b}	38,15±11,50 ^a	43,91±14,78 ^b	<0,001
DL _{CO} /VA (%)	81,49±20,33 ^a	66,46±18,81 ^a	72,03±24,23	0,005
FVC/DL _{CO} (%)	1,66±0,55	1,93±0,57	2,00±0,85	0,044
sPAP (mmHg)	55,24±16,72 ^{a, b}	72,64±19,79 ^a	82,76±20,02 ^b	<0,001
mPAP (mmHg)	30,22±7,64 ^{a, b}	42,50±9,67 ^{b, c}	48,72±11,80 ^c	<0,001
PCWP (mmHg)	15,09±6,80	16,35±6,84 ^c	12,24±4,63 ^c	0,017
CI (Lmin ⁻¹ mt ⁻²)	3,23±1,45 ^{a, b}	2,33±0,43 ^a	2,32±0,60 ^b	<0,001
RAP (mmHg)	9,56±4,95 ^a	15,29±8,06 ^{a, c}	10,55±3,95 ^c	<0,001
PVR (Wu)	3,17±1,54 ^{a, b}	6,59±2,96 ^{a, c}	8,61±4,51 ^{b, c}	<0,001
TPG (mmHg)	15,32±7,64 ^{a, b}	25,95±12,21 ^{a, c}	36,32±11,24 ^{b, c}	<0,001
6MWT (mt)	318,38±130,34 ^a	219,09±79,82 ^a	275,95±115,40	0,001
WHO functional class (n)	2,61±0,84 ^{a, b}	3,18±0,49 ^a	3,22±0,58 ^b	<0,001
Number of drugs (n)	0,71±0,79 ^b	1,08±0,81 ^c	2,14±0,71 ^{b, c}	<0,001
Exitus (%)	33% ^a	75% ^{a, c}	38% ^c	<0,001
Months of survival (n)	49,58±18,46 ^a	24,10±22,76 ^{a, c}	52,78±13,81 ^c	<0,001

Cluster 1 had a mean age of 68.57 ± 10.54 years, significantly different from Cluster 2's 71.36 ± 8.32 years and Cluster 3's 61.11 ± 13.50 years, with Cluster 3 being the youngest ($p < 0.001$). Significant differences were also observed in respiratory function among the clusters, with Cluster 2 exhibiting significantly lower FVC and FEV₁ compared to the other clusters, particularly FVC 68.01 ± 12.66 ($p < 0.001$) and FEV₁ 68.12 ± 10.20 ($p < 0.015$).

Furthermore, patients in Cluster 2 had lower DL_{CO} ($p < 0,001$) and DL_{CO} /V_A values ($p = 0,005$) compared to those in Cluster 1, while Cluster 3 showed intermediate values between the two, with statistically significant differences.

Pulmonary arterial pressure measurements revealed that patients in Clusters 2 and 3 had higher sPAP and mPAP values compared to Cluster 1 ($p < 0,001$), with Cluster 3 showing the highest values in both measurements (sPAP 82.76 ± 20.02 and mPAP 48.72 ± 11.80 ; $p < 0,001$). Additionally, Cluster 2 had higher PAdx, RVP, and GTP values compared to the other groups ($p < 0,001$).

Regarding functional capacity, Cluster 2 had a significantly lower 6MWT distance (219.09 ± 79.82 meters) compared to the other clusters (Cluster 1 = $318,38 \pm 130,34$ meters, Cluster 3 = $275,95 \pm 115,40$ meters; $p = 0,001$), while the number of prescribed medications was significantly higher in Cluster 3 (Cluster 3 = 2.14 ± 0.71 , Cluster 1 = $0,71 \pm 0,79$, Cluster 2 = $1,08 \pm 0,81$; $p < 0,001$). Finally, the mortality rate was higher in Cluster 2 (75%) compared to Cluster 1 (33%) and Cluster 3 (38%). These mortality differences were confirmed by Kaplan-Meier curves (**Figure 2**).

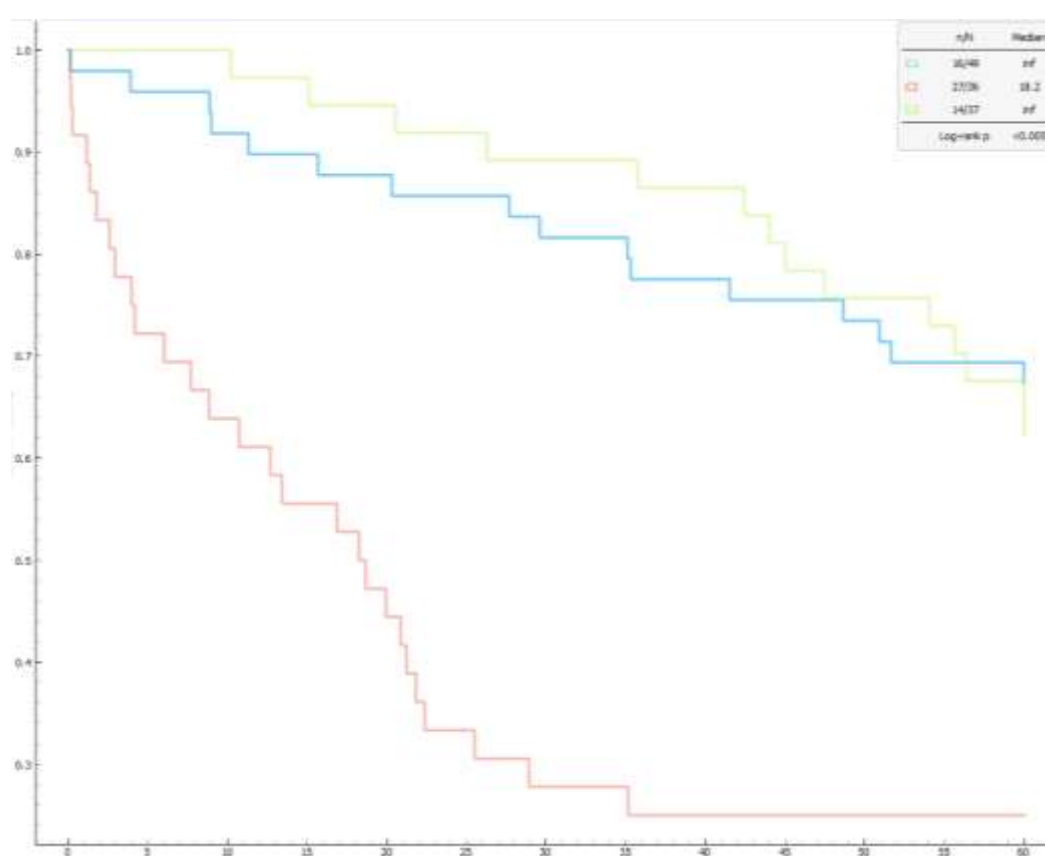


Figure 2. Kaplan-Meier survival curves for clusters. The x-axis represents 5-year survival in months (from 0 to 60 months), while the y-axis shows the probability of survival (1 indicates 100% survival). The legend indicates theHumbert M, Kovacs G, Hoeper MM, Badagliacca R, Berger RMF, Brida M, Carlsen J, Coats AJS, Escribano-Subias P, Ferrari P, Ferreira DS, Ghofrani HA, Giannakoulas G, Kiely DG, Mayer E, Meszaros G, Nagavci B, Olsson KM, Pepke-Zaba J, Quint JK, Rådegran G, Simonneau G, Sitbon O, Tonia T, Toshner M, Vachiery JL, Vonk Noordegraaf A, Delcroix M, Rosenkranz S; ESC/ERS Scientific Document Group. 2022 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension. Eur Respir J. 2023 Jan 6;61(1):2200879. number of deaths out of the total number of patients in each cluster (n/N). The median represents the median survival time. The median is "inf" (infinite) for clusters 1 and 3, meaning that more than half of the patients survived beyond the observation period. C1 (in red) has the lowest survival probability, C2 (in blue) has an intermediate survival probability, and C3 has the highest survival probability.

A Logistic Regression predictive model identified three risk factors for mortality: age, the number of medications taken for PH, and NYHA class, with an accuracy of 74.4%.

4. Discussion

PH is a life-threatening condition associated with increased mortality regardless of the classification and underlying etiology [15,16].

In this regard, it seems strategically important to have optimal, multidisciplinary and integrated management of this condition [17] in order to diagnose it early, treat it appropriately and improve outcomes [18].

In this context, there is a strict need for risk stratification algorithms to define the prognosis of these patients and guide diagnostic-therapeutic decisions.

This study aimed to identify homogeneous patient groups (clusters) within the examined population, analyze their mortality rates, and develop a five-year mortality prediction model for patients with PH using ML techniques.

Through clustering, three distinct phenotypes with varying clinical and demographic characteristics were identified.

Cluster 1: predominantly female (57%), with 69% belonging to clinical groups other than Group 3 (according to the 2022 ESC/ERS classification), characterized by relatively favorable clinical parameters, intermediate age (68.57 ± 10.54 years), and better cardiac and respiratory function.

Cluster 2: Composed of 50% women, with the highest average age (71.36 ± 8.32 years). This cluster includes 44% of patients in Group 3, with worse respiratory function and intermediate cardiac function.

Cluster 3: The youngest group (61.11 ± 13.50 years), mostly male (65%), primarily from non-respiratory PH groups (81%), showing intermediate respiratory dysfunction but the poorest cardiac function.

Mortality analysis revealed that Cluster 2 had the lowest survival rates, with a 75% mortality rate, significantly higher than Cluster 1 (33%) and Cluster 3 (38%). This evidence implies two considerations: *in primis*, more severe pulmonary and cardiac conditions are linked to worse outcomes, emphasizing the importance of integrated assessment in PH management; *in secundis*, elderly patients with worse lung function are often associated with higher mortality in PH due to the combined effects of aging and lung deterioration.

The effects of aging on the respiratory system have been widely studied [19] and result from a miscellany of progressive decline in lung function (secondary to altered respiratory mechanics [20]) and loss of total capillary volume.

On the other hand, age-related vascular stiffening has also been shown to involve the pulmonary vascular bed, with age-related increase of systolic pulmonary artery pressure (sPAP) of about 1 mmHg per decade [21].

Decline in spirometric data, including FEV1 and FVC, predicted was independently associated with an increased risk of death [22].

This is in agreement with our study: Cluster 1, which has the best cardiac and respiratory function values, has the lowest mortality.

Younger patients with severe cardiac dysfunction often show better survival rates than older patients with intermediate cardiac dysfunction (Cluster 3, younger but associated with worse cardiac function, has lower mortality than Cluster 2, consisting of older but intermediate cardiac function patients), probably due to greater physiologic resilience and better response to treatments [23].

The study offers important insights into managing PH by identifying distinct patient clusters and developing a predictive model for 5-year mortality. These clusters underscore the heterogeneity within the PH population, highlighting the need for tailored treatment. A ML model identified key predictors of mortality (age, medications taken for PH, NYHA class) showing that older patients in advanced NYHA classes face higher mortality.

Numerous studies have confirmed that older patients have a higher comorbidity and a reduced ability to recover as well as NYHA functional class is a strong predictor of mortality, in fact patients in more advanced NYHA classes (III and IV) show a significantly lower probability of survival than those in classes I and II.

This evidence agrees with data from the REVEAL registry, which identified predictors of 1-year mortality in patients with PH, finding that advanced age was significantly associated with increased mortality, as well as worse NYHA functional class and more treatment medications [24].

Similar evidence was reached by the COMPERA registry, which confirmed that advanced age is associated with increased mortality risk in patients with PAH [25].

This model provides simplicity and accessibility by focusing on just three easily obtainable variables, offering clinicians a practical tool for assessing patient prognosis and tailoring therapeutic strategies.

Although widely used, existing risk models such as REVEAL (Registry to Evaluate Early and Long-Term Pulmonary Arterial Hypertension Disease Management) [26], COMPERA (Comparative, Prospective Registry of Newly Initiated Therapies for Pulmonary Hypertension) [27], SPAHR (The Swedish Pulmonary Arterial Hypertension Register) [28] and FPHR (French PH registry) [29] scores have limitations, including complexity and reliance on invasive parameters. Other risk-stratification tools have been developed from the US REVEAL (*i.e.* REVEAL 2.0 risk score calculator and REVEAL Lite 2) [30,31].

The ML model developed in this study offers significant advantages in terms of simplicity, accessibility, and consistent performance across diverse patient populations.

This is not the first study to apply AI to PH. Previous research has employed AI to quantify the extent of pulmonary fibrosis on CT scans, showing a strong correlation with increased mortality [32]. AI has also been used to develop algorithms for detecting PH through chest X-ray images [33]. Furthermore, other studies have explored AI models to predict elevated pulmonary pressures using data from echocardiograms [34] or electrocardiograms [35].

The strength of our model lies in its simplicity, as it is based on only a few parameters. Despite this, it provides a valuable method for phenotyping, helping us identify high-risk patient clusters that may benefit from a more aggressive therapeutic approach.

Limitations

This study has several limitations. The sample size, while adequate, limits the generalizability of the findings. Additionally, some variables that could influence mortality, such as genetic factors and comorbidities, were not included in the analysis.

5. Conclusions

This study identified patient clusters and developed a predictive model for 5-year mortality in PH patients, enhancing the understanding of PH progression and informing clinical decision-making. The integration of traditional statistical methods with ML techniques demonstrates the potential to improve diagnostic accuracy and patient outcomes in PH. Future research should focus on validating these findings in larger populations and exploring additional variables that may impact mortality.

Continuous data are expressed as mean±standard deviation, while categorical as percentage. Abbreviations. PH: pulmonary hypertension; FVC: forced vital capacity, FEV₁: forced expiratory volume in 1 s; DL_{CO}: diffusing capacity of the lung for carbon monoxide; V_A: alveolar volume; sPAP: systolic pulmonary artery pressure; mPAP: mean pulmonary artery pressure; PCWP: pulmonary capillary wedge pressure; CI: cardiac index; RAP: mean right atrial pressure; PVR (WU): pulmonary vascular resistance (Wood units); TPG: transpulmonary gradient; 6MWT: six-minute walking test; WHO: World Health Organization.

Continuous data are expressed as mean±standard deviation, while categorical as percentage. Significant differences in Tukey's post-hoc tests are indicated as follows: a for C1 vs C2, b for C1 vs C3, c for C2 vs C3. Abbreviations. PH: pulmonary hypertension; FVC: forced vital capacity, FEV₁: forced expiratory volume in 1 s; DL_{CO}: diffusing capacity of the lung for carbon monoxide; V_A: alveolar volume; sPAP: systolic pulmonary artery pressure; mPAP: mean pulmonary artery pressure;

PCWP: pulmonary capillary wedge pressure; CI: cardiac index; RAP: mean right atrial pressure; PVR (Wu): pulmonary vascular resistance (wood units); TPG: transpulmonary gradient; 6MWT: six-minute walking test; WHO: World Health Organization.

The Receiver Operating Characteristic (ROC) curve shows the relationship between the true positive rate on the y-axis (TPR, sensitivity) and the false positive rate on the x-axis (FPR, 1-specificity).

The Kaplan-Meier graph shows significant differences in survival among the three clusters of pulmonary hypertension patients. The log-rank test confirms that the differences in survival between clusters are statistically significant ($p < 0.005$).

References

1. Pepke-Zaba J, Quint JK, Rådegran G, Simonneau G, Sitbon O, Tonia T, Toshner M, Vachiery JL, Vonk Noordegraaf A, Delcroix M, Rosenkranz S; ESC/ERS Scientific Document Group. 2022 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension. *Eur Respir J*. 2023 Jan 6;61(1):2200879.
2. Kovacs G, Bartolome S, Denton CP, et al. Definition, classification and diagnosis of pulmonary hypertension. *Eur Respir J* 2024; 64: 2401324 [DOI: 10.1183/13993003.01324-2024]
3. Rose-Jones LJ and McLaughlin VV. Pulmonary hypertension: types and treatments. *Curr Cardiol Rev* 2015; 11: 73–79.
4. Coghlan JG, Denton CP, Grünig E, Bonderman D, Distler O, Khanna D, Müller-Ladner U, Pope JE, Vonk MC, Doelberg M, Chadha-Boreham H, Heinzl H, Rosenberg DM, McLaughlin VV, Seibold JR; DETECT study group. Evidence-based detection of pulmonary arterial hypertension in systemic sclerosis: the DETECT study. *Ann Rheum Dis*. 2014 Jul;73(7):1340-9.
5. Boon GJAM, Ende-Verhaar YM, Bavalua R, El Bouazzaoui LH, Delcroix M, Dzikowska-Diduch O, Huisman MV, Kurnicka K, Mairuhu ATA, Middeldorp S, Pruszczyk P, Ruigrok D, Verhamme P, Vliegen HW, Vonk Noordegraaf A, Vriend JWJ, Klok FA; InShape II study group. Non-invasive early exclusion of chronic thromboembolic pulmonary hypertension after acute pulmonary embolism: the InShape II study. *Thorax*. 2021 Oct;76(10):1002-1009.
6. Siontis KC, Noseworthy PA, Attia ZI, Friedman PA. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. *Nat Rev Cardiol*. 2021 Jul;18(7):465-478.
7. Tison GH, Zhang J, Delling FN, Deo RC. Automated and Interpretable Patient ECG Profiles for Disease Detection, Tracking, and Discovery. *Circ Cardiovasc Qual Outcomes*. 2019 Sep;12(9):e005289.
8. Bergemann R, Allsopp J, Jenner H, Daniels FA, Drage E, Samyshkin Y, Schmitt C, Wood S, Kiely DG, Lawrie A; SPHInX Project team. High levels of healthcare utilization prior to diagnosis in idiopathic pulmonary arterial hypertension support the feasibility of an early diagnosis algorithm: the SPHInX project. *Pulm Circ*. 2018 Oct-Dec;8(4):2045894018798613.
9. Kiely DG, Doyle O, Drage E, Jenner H, Salvatelli V, Daniels FA, Rigg J, Schmitt C, Samyshkin Y, Lawrie A, Bergemann R. Utilising artificial intelligence to determine patients at risk of a rare disease: idiopathic pulmonary arterial hypertension. *Pulm Circ*. 2019 Nov 20;9(4):2045894019890549.
10. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, Crapo R, Enright P, van der Grinten CP, Gustafsson P, Jensen R, Johnson DC, MacIntyre N, McKay R, Navajas D, Pedersen OF, Pellegrino R, Viegi G, Wanger J; ATS/ERS Task Force. Standardisation of spirometry. *Eur Respir J*. 2005 Aug;26(2):319-38.
11. Graham BL, Brusasco V, Burgos F, Cooper BG, Jensen R, Kendrick A, MacIntyre NR, Thompson BR, Wanger J. 2017 ERS/ATS standards for single-breath carbon monoxide uptake in the lung. *Eur Respir J*. 2017 Jan 3;49(1):1600016.
12. ATS Committee on Proficiency Standards for Clinical Pulmonary Function Laboratories. ATS statement: guidelines for the six-minute walk test. *Am J Respir Crit Care Med*. 2002 Jul 1;166(1):111-7.
13. SPSS Software | IBM. <https://www.ibm.com/spss> (2024).
14. Orange Data Mining. <https://orangedatamining.com/>.
15. Hoepfer MM, Humbert M, Souza R, et al. A global view of pulmonary hypertension. *Lancet Respir Med*. 2016;4(4):306-322

- Stacy A. Mandras, Hirsch S. Mehta, Anjali Vaidya, Pulmonary Hypertension: A Brief Guide for Clinicians, 16. Mayo Clinic Proceedings, Volume 95, Issue 9, 2020, Pages 1978-1988, ISSN 0025-6196.
17. Osterberg L, Blaschke T. Adherence to medication. *N Engl J Med* 2005;353: 487–497.
18. Early detection and management of pulmonary arterial hypertension. Marc Humbert J. Gerry Coghlan Dinesh Khanna *European Respiratory Review* 2012 21(126): 306-312.
- Janssens JP, Pache JC, Nicod LP. Physiological changes in respiratory function associated with ageing. *Eur Respir J* 1999; 13: 197–205.
19. Knudson RJ, Slatin RC, Lebowitz MD, et al. The maximal expiratory flow±volume curve: normal standards, 20. variability, effects of age. *Am Rev Respir Dis* 1976; 113: 587–599.
21. Pulmonary hypertension in the elderly: a different disease? Grégory Berra Stéphane Noble Paola M. Soccia Show More *Breathe* 2016 12(1): 43-49.
22. Pescatore J, Bittner M, D'Alonzo G, Weaver S, Gayen S. Predictors of Mortality in Pulmonary Hypertension-Associated Chronic Lung Disease. *J Clin Med*. 2024 Jun 14;13(12):3472. doi: 10.3390/jcm13123472. PMID: 38929999; PMCID: PMC11205208.
23. Guazzi M, Galiè N. Pulmonary hypertension in left heart disease. *Eur Respir Rev*. 2012 Dec 1;21(126):338-46. doi: 10.1183/09059180.00004612. PMID: 23204122; PMCID: PMC9487233.
24. Benza, Raymond L. et al. The REVEAL Registry Risk Score Calculator in Patients Newly Diagnosed With Pulmonary Arterial Hypertension *CHEST*, Volume 141, Issue 2, 354 – 362.
25. Hoeper, M. M. et al. Idiopathic pulmonary arterial hypertension phenotypes determined by cluster analysis from the COMPERA registry. *J. Heart Lung Transplant*. 39, 1435–1444 (2020).
26. Benza RL, Kanwar MK, Raina A, Scott JV, Zhao CL, Selej M, Elliott CG, Farber HW. Development and Validation of an Abridged Version of the REVEAL 2.0 Risk Score Calculator, REVEAL Lite 2, for Use in Patients With Pulmonary Arterial Hypertension. *Chest*. 2021 Jan;159(1):337-346.
27. Hoeper MM, Kramer T, Pan Z, Eichstaedt CA, Spiesshoefer J, Benjamin N, Olsson KM, Meyer K, Vizza CD, Vonk-Noordegraaf A, Distler O, Opitz C, Gibbs JSR, Delcroix M, Ghofrani HA, Huscher D, Pittrow D, Rosenkranz S, Grünig E. Mortality in pulmonary arterial hypertension: prediction by the 2015 European pulmonary hypertension guidelines risk stratification model. *Eur Respir J*. 2017 Aug 3;50(2):1700740.
28. Kylhammar D, Kjellström B, Hjalmarsson C, Jansson K, Nisell M, Söderberg S, Wikström G, Rådegran G. A comprehensive risk stratification at early follow-up determines prognosis in pulmonary arterial hypertension. *Eur Heart J*. 2018 Dec 14;39(47):4175-4181.
29. Boucly A, Weatherald J, Savale L, Jaïs X, Cottin V, Prevot G, Picard F, de Groote P, Jevnikar M, Bergot E, Chaouat A, Chabanne C, Bourdin A, Parent F, Montani D, Simonneau G, Humbert M, Sitbon O. Risk assessment, prognosis and guideline implementation in pulmonary arterial hypertension. *Eur Respir J*. 2017 Aug 3;50(2):1700889.
30. Benza RL, Gomberg-Maitland M, Elliott CG, Farber HW, Foreman AJ, Frost AE, et al. Predicting survival in patients with pulmonary arterial hypertension: the REVEAL risk score calculator 2.0 and comparison with ESC/ERS-based risk assessment strategies. *Chest* 2019;156:323–337.
31. Benza RL, Kanwar MK, Raina A, Scott JV, Zhao CL, Selej M, et al. Development and validation of an abridged version of the REVEAL 2.0 risk score calculator, REVEAL Lite 2, for use in patients with pulmonary arterial hypertension. *Chest* 2021;159: 337–346.
32. Dwivedi K, Sharkey M, Delaney L, Alabed S, Rajaram S, Hill C, Johns C, Rothman A, Mamalakis M, Thompson AAR, Wild J, Condliffe R, Kiely DG, Swift AJ. Improving Prognostication in Pulmonary Hypertension Using AI-quantified Fibrosis and Radiologic Severity Scoring at Baseline CT. *Radiology*. 2024 Feb;310(2):e231718.
33. Imai S, Sakao S, Nagata J, Naito A, Sekine A, Sugiura T, Shigeta A, Nishiyama A, Yokota H, Shimizu N, Sugawara T, Nomi T, Honda S, Ogaki K, Tanabe N, Baba T, Suzuki T. Artificial intelligence-based model for predicting pulmonary arterial hypertension on chest x-ray images. *BMC Pulm Med*. 2024 Feb 27;24(1):101.
34. Hirata Y, Tsuji T, Kotoku J, Sata M, Kusunose K. Echocardiographic artificial intelligence for pulmonary hypertension classification. *Heart*. 2024 Mar 22;110(8):586-593.

35. Liu CM, Shih ESC, Chen JY, Huang CH, Wu IC, Chen PF, Higa S, Yagi N, Hu YF, Hwang MJ, Chen SA. Artificial Intelligence-Enabled Electrocardiogram Improves the Diagnosis and Prediction of Mortality in Patients With Pulmonary Hypertension. JACC Asia. 2022 May 17;2(3):258-270.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.