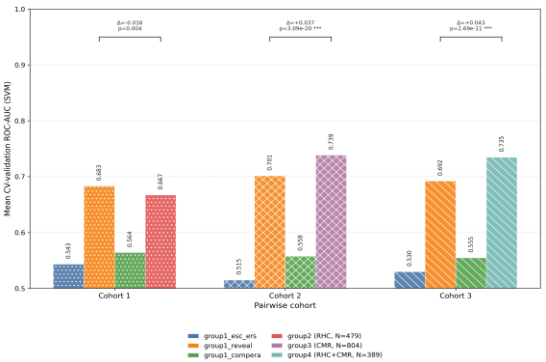
Machine-Learning Models for Improved 1-Year Mortality Prediction in Newly Diagnosed Pulmonary Arterial Hypertension Patients

Ning Wang (1), Ze Ming Goh (1), Harry Saxton (4), Marcelle Paula-Ribeiro (1), Joe Longbottom (1), Maximilian Balmus (2), Chris Lanyon (3), Tiffany Baptiste (2), Alexander Rothman (1), Martin Wilkins (2), Richard Wilkinson (3), Keith Worden (1), David Wagg (1), Steven Niederer (2), Xu Xu (1), Richard Clayton (1)

1. University of Sheffield, UK; 2. Imperial College London, UK; 3. University of Nottingham, UK; 4. Astellas Pharma, UK

Introduction: Predicting 1-year mortality in newly diagnosed pulmonary arterial hypertension (PAH) is important because the first year after diagnosis is critical for treatment planning and transplant assessment. This study evaluated whether multimodal machine-learning models could improve prediction beyond established risk approaches.

Methods: This retrospective study used ASPIRE registry data from 2008 to 2025. 1-year mortality was modelled using demographic, laboratory, right heart catheterisation (RHC), and cardiovascular magnetic resonance (CMR) data. REVEAL, ESC/ERS, COMPERA, and WHO functional class were benchmarked using 4-fold cross-validation. Models were developed using RHC-only, CMR-only, and combined RHC+CMR features. Class imbalance was addressed using SMOTE-NC and class weighting. Backward feature elimination derived reduced models.



Pairwise ROC-AUC comparison using SVM.

Results: REVEAL performed best among established risk approaches (AUC=0.699). The support vector machine (SVM) with SMOTE-NC performed best overall. CMR-only and combined RHC+CMR models outperformed matched REVEAL benchmarks (0.739 vs 0.701, Δ=0.037, $p<0.001$; and 0.735 vs 0.692, Δ=0.043, $p<0.001$), whereas the RHC-only model performed slightly worse (0.667 vs 0.683, Δ=-0.016, p=0.004). A reduced 9-feature combined model achieved the highest cross-validation AUC of 0.797 (SD=0.075).