

Fenotipos clínicos en síndrome coronario agudo mediante machine learning-based clustering sobre PCA continuo: cohorte retrospectiva 2016–2025

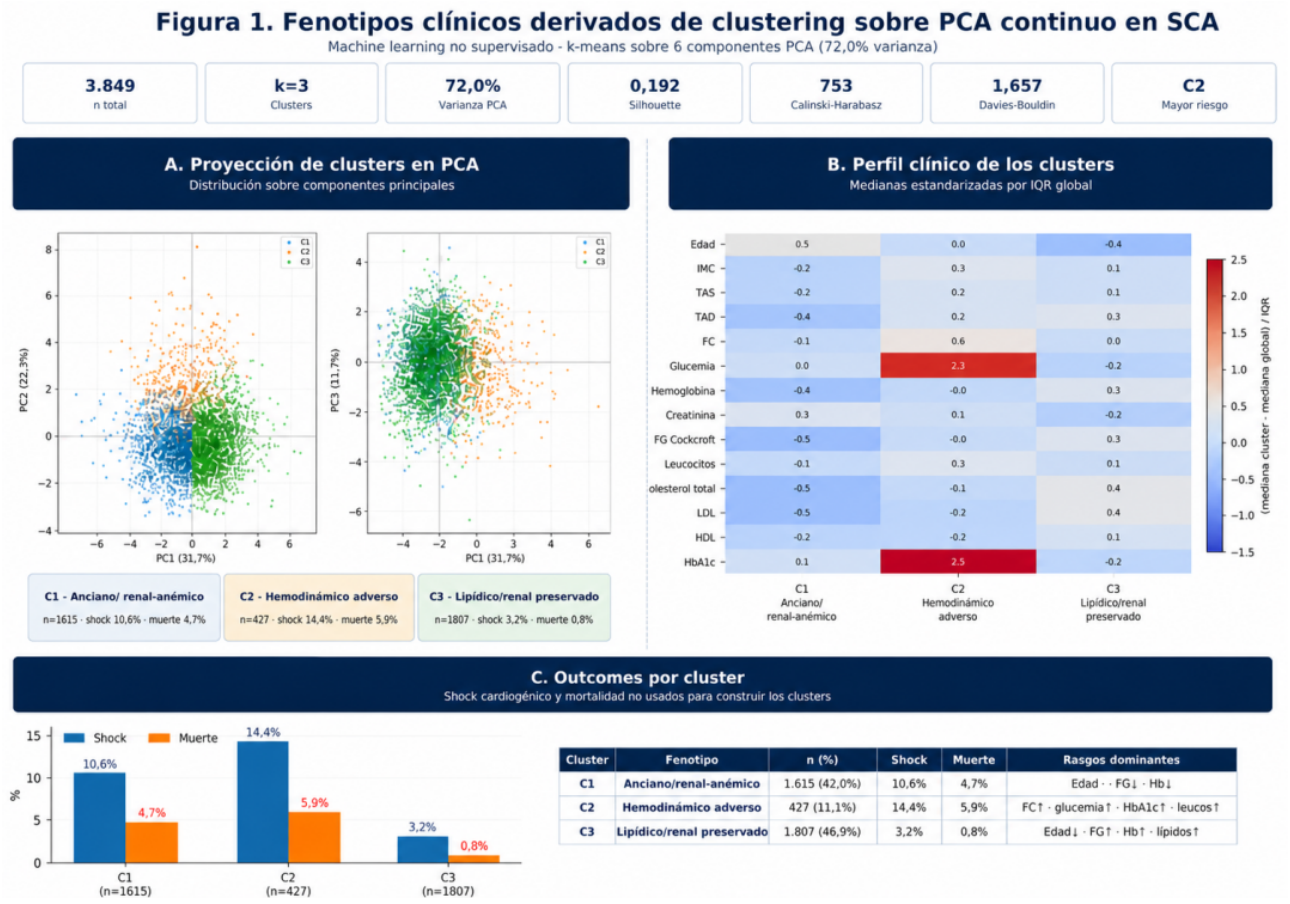
Sergi Felis Badia(1), Álvaro Jara Díaz(1), Arnau Cabello Ramírez(1), Esther Papiol Morales(1), Mireia Martínez Rubió(1), Alba Bermúdez-Jiménez(1), Alba Orellana Morote(1), Laura Botella Pascual(1), Tania Ramírez Martínez(1), Daniele Luiso(1), Luis Enrique Martín Alfaro(1), M. Isabel Hernández Martín(1), Emilia Blanco Ponce(1), Gemma Mateus Porta(1) y Juan Manuel Casanova Sandoval(1) de (1)Servei de Cardiologia. Hospital Universitari Arnau de Vilanova, Lleida - España

Introducción: El síndrome coronario agudo presenta gran heterogeneidad clínica y pronóstica. Las técnicas de machine learning no supervisado pueden identificar fenotipos clínicos latentes a partir de variables continuas basales. El objetivo fue aplicar clustering sobre componentes principales (PCA), para identificar fenotipos clínico-analíticos en pacientes con síndrome coronario agudo y evaluar su asociación posterior con shock cardiogénico y mortalidad.

Métodos: Estudio observacional retrospectivo de pacientes ingresados por síndrome coronario agudo entre 2016 y 2025. Se realizó PCA sobre variables continuas basales y se utilizaron los primeros seis componentes, que explicaron el 72,0% de la varianza, como entrada para clustering. Se compararon soluciones k-means y jerárquicas entre k=2 y k=7. Se mantuvo k-means con tres clusters por equilibrio entre métricas internas, tamaño de grupos e interpretabilidad clínica. Las métricas del modelo fueron: silhouette 0,192, Calinski-Harabasz 753 y Davies-Bouldin 1,657. Shock cardiogénico y mortalidad no se utilizaron para generar los clusters, sino que se analizaron posteriormente como desenlaces.

Resultados: Se identificaron tres fenotipos. C1, anciano/renal-anémico, incluyó 1.615 pacientes, 42,0%, con mayor edad, menor filtrado glomerular, mayor creatinina y menor hemoglobina; presentó shock en 10,6% y mortalidad en 4,7%. C2, hemodinámico adverso, incluyó 427 pacientes, 11,1%, con mayor frecuencia cardíaca, hiperglucemia marcada, HbA1c elevada y leucocitosis; mostró el mayor riesgo, con shock en 14,4% y mortalidad en 5,9%. C3, lipídico/renal preservado, incluyó 1.807 pacientes, 46,9%, con menor edad, mejor función renal, mayor hemoglobina y perfil lipídico más alto; presentó el menor riesgo, con shock en 3,2% y mortalidad en 0,8%. Las diferencias entre clusters fueron significativas para shock y mortalidad, $p<0,001$.

Conclusiones: En esta cohorte retrospectiva 2016–2025, el machine learning-based clustering sobre PCA continuo identificó tres fenotipos clínicos diferenciados y asociados a distinto riesgo de shock cardiogénico y mortalidad. El fenotipo hemodinámico adverso concentró el mayor riesgo, mientras que el fenotipo lipídico/renal preservado mostró el mejor pronóstico. Este enfoque exploratorio puede complementar la estratificación convencional, aunque requiere validación temporal o externa.



Cluster	Fenotipo	n (%)	Shock	Muerte	Rasgos dominantes
C1	Anciano/renal-anémico	1.615 (42,0%)	10,6%	4,7%	Edad · · FG↓ · Hb↓
C2	Hemodinámico adverso	427 (11,1%)	14,4%	5,9%	FC↑ · glucemia↑ · HbA1c↑ · leucos↑
C3	Lipídico/renal preservado	1.807 (46,9%)	3,2%	0,8%	Edad↓ · FG↑ · Hb↑ · lípidos↑

Outcomes por clúster

Clúster	Fenotipo	n (%)	Shock	Muerte	Rasgos dominantes
C1	Anciano/renal-anémico	1615 (42.00%)	10.6%	4.7%	Edad elevada - FG bajo - Hb baja
C2	Hemodinámico adverso	427 (11.1%)	14.4%	5.9%	FC elevada - HbA1c elevada - Leucos elevados
C3	Lipídico/renal preservado	1807 (46.9%)	3.2%	0.8%	Edad baja - FG alto - Hb alta - lipidos elevados

FG: filtrado glomerular, Hb: hemoglobina, FC: frecuencia cardíaca, HbA1c: hemoglobina glicosilada (A1c), Leucos: leucocitos