AI for bariatric surgery patient differentation

MULTIVARIABLE ANALYSIS AND MACHINE LEARNING APPROACH TO CLASSIFY PRE-BARIATRIC SURGERY PATIENTS TO ESTABLISH A POSTOPERATIVE SURVEILLANCE SYSTEM IN PRIMARY CARE

Clínica Puerto Montt
achs salud

Grecia Estefania Rojas Rojas.
Gustavo Martinez.

INTA
UNIVERSIDAD DE CHILE

Francisco Perez.
Christian Hodar.

No conflicts of interest.

General objective

How similar are the patients?

65 features or variables

To identify **preoperative patterns** in bariatric surgery patients through

Multivariate analysis and **Machine learning** using demographic, clinical, and paraclinical data

Establish groups that serve as a **baseline**

for the subsequent development of a postoperative monitoring system in primary care.

Investigation in progress

1

Samples

130 adults

Clínica ACHS Salud Puerto Montt
Chile

All chilean adults admitted with obesity and comorbilities in evaluation before bariatric surgery

112 Women 18 Men

Median: 37 years old (31-44)

2

Data management

- Manual extraction of medical records and printed tests by a member of the clinical team.
- Masked data and consolidation of this data in an Excel matrix.
- Standardization of variable names and construction of frequency tables.
- Filtering of variables with low frequency (>125 zeros).
- Validation of variables by expected domains, analysis by histograms, and correlations.
- Total 65 variables

3

Missing data

Multiple Imputation by Chained Equations

4

Relationship and clustering

Multivariate analysis

Supervised vs. unsupervised (AI)

Machine learning techniques

Hierarchical clustering

Non-hierarchical clustering

Support Vector Machine

Multilayer Perceptron

5

Results preliminars

To establish groups that serve as a baseline for the subsequent development of a postoperative monitoring system

Multivariate analysis

Database ✳ → PCA mix → **Components (*k*)** → + 3 groups*

Unsupervised

Hierarchical clustering

Linkage Ward+ Gower distance (A)

Dendrogram:
3 groups*

Average values*

Non-hierarchical clustering

K-means (B)

Average values*

K-medians (C)

Average values*

Each observation was assigned to a cluster in each method (A, B, C).

Validate grouping quality:

Silhoutte Index

There was no clear separation

Confusions Matrix

Consensus-co-asociation (A+B+C)= average → Adjusted Rand Index (ARI): **Consensus** vr Method A - B -C → **Concordance**

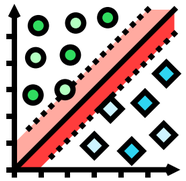
Linkage Ward+ Gower distance (A)

0.739 (alta similitud)

Supervised

Support Vector Machine

3 groups



Learn to reproduce the **Consensus**

Machine learning techniques

Multilayer Perceptron

3 groups

Build baseline

Linkage Ward+ Gower distance (A) ✳

Dendrogram: 3 groups*

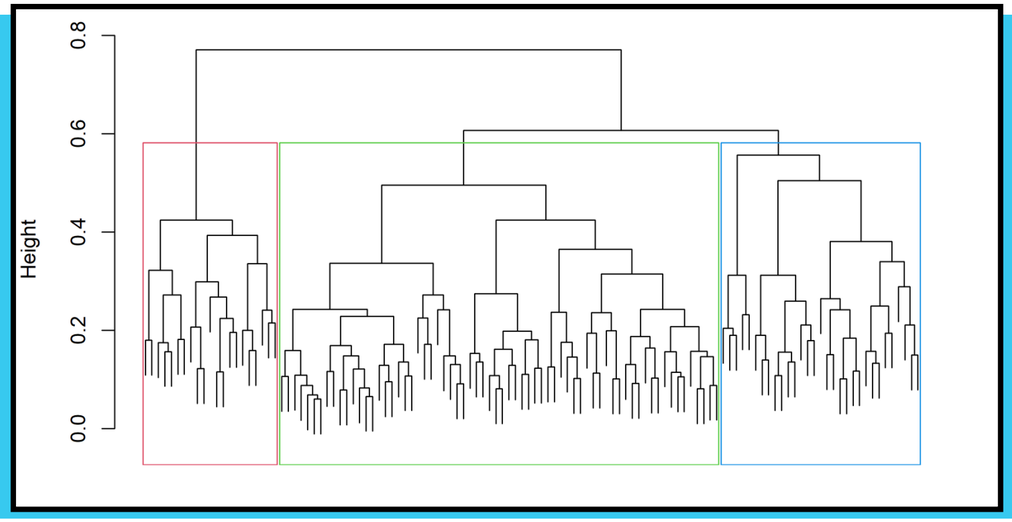
Multilayer Perceptron

Support Vector Machine

3 groups

Relationship and clustering

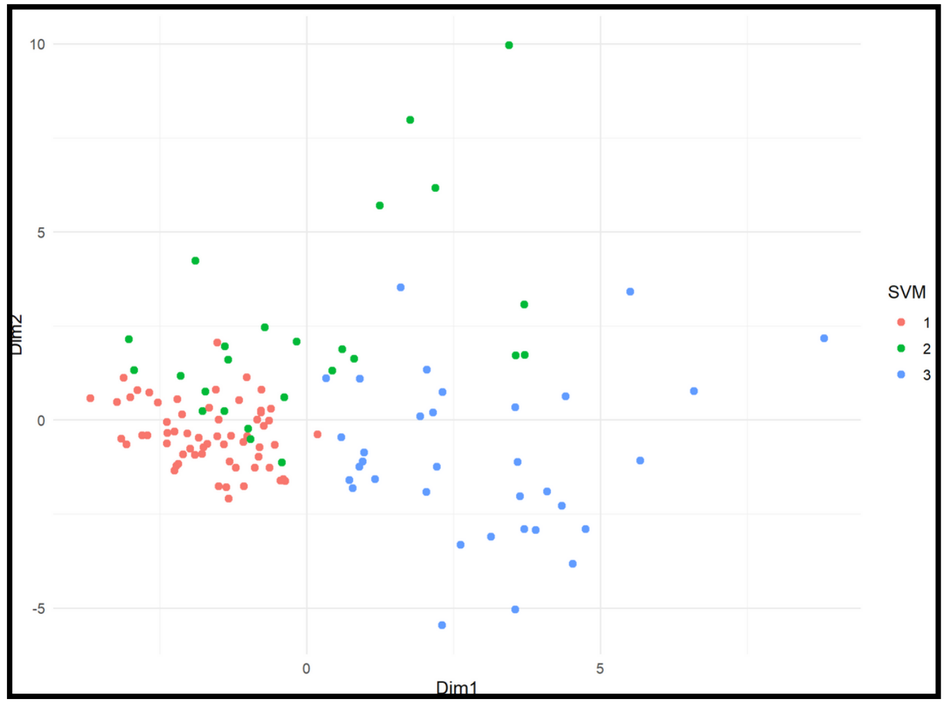
Unsupervised



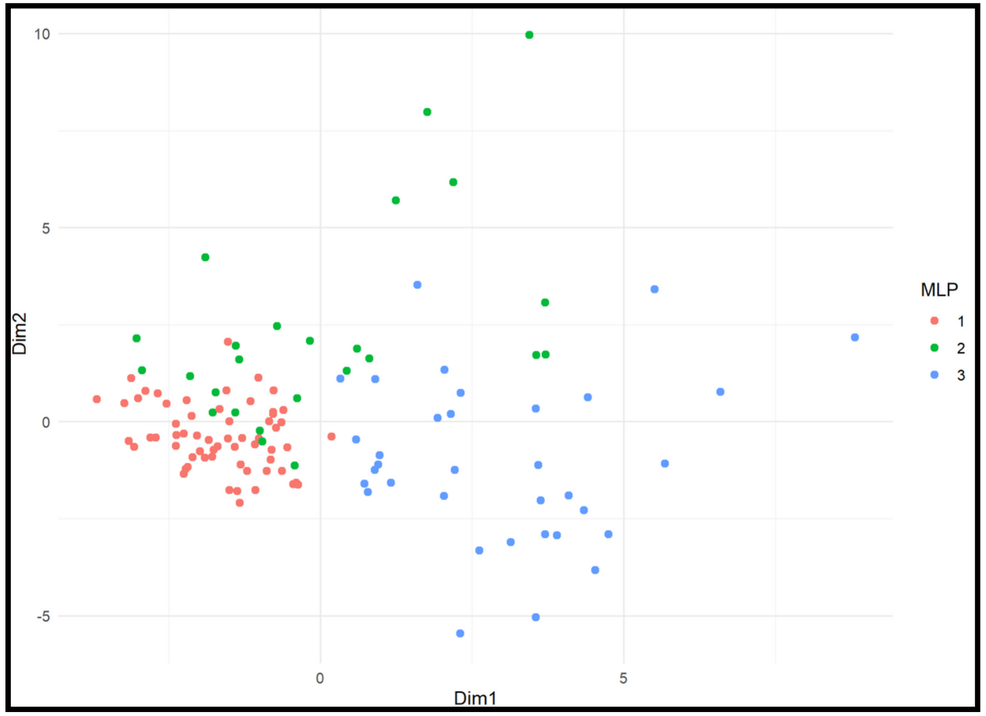
Dendrogram

Linkage Ward+ Gower distance (A)

Supervised



Support Vector Machine



Multilayer Perceptron

Preliminary results

Group 1 : **“Metabolic Balance Family”**

Profile suggestive of patients with better metabolic control, low blood glucose, and lower insulin resistance.

Values lower than the global average in variables: vitamin B12, ferritin, triglycerides, blood glucose, GGT, basal insulin, and Homa.

HDL appears slightly higher.

Group 2 : **“Dysglycemia–Hypertriglyceridemia Family”**

This group has a higher concentration of hyperglycemia and an unfavorable lipid profile.

Ferritin, B12, iron, and weight appear lower..

Group 3 : **“Severe Insulin Resistance–Hepatic Dysfunction Family”**

Ferritin and B12 are well above average.

Also higher: triglycerides, GPT, weight, basal insulin, GGT.

HDL appears lower.

Results, implication and gaps.

1. **Similitude of results using** Database = PCA mix (reduced dimensionality).
2. **Patients categorized in 3 clusters**
3. **Statistical and machine learning** analyses produced the 3 clusters.

1. To check for **convenience sampling.**
2. To **increase** the number of patients (N).
3. To **monitor post-operate exams to evaluate** model prediction

1. Future patients should be allocated to one of these families.
2. Generate differentiation plans to post-operate managments.

In the future, we hope to have a complete database with data from preoperative and postoperative evaluations.

Clínica
Puerto Montt

achs salud

INTA
UNIVERSIDAD DE CHILE

Thank You

Grecia Estefania Rojas Rojas.
greciasalud@gmail.com

A continuación: Diapositivas de reserva

FF029 / Abs. ID 83

Topic: A543. AI (artificial intelligence), virtual reality in endoscopy, digital operating room

FREE PAPERS 02: HERNIA/VTE/ARTIFICIAL INTELLIGENCE

MULTIVARIABLE ANALYSIS AND MACHINE LEARNING APPROACH TO CLASSIFY PRE-METABOLIC BARIATRIC SURGERY PATIENTS TO ESTABLISH A POSTOPERATIVE SURVEILLANCE SYSTEM IN PRIMARY CARE

Grecia Rojas[1], Francisco Pérez-Bravo[2], Christian Hodar[3], Gustavo Martínez[1]
[1]Clinica Puerto Montt, Centro De Manejo Integral De La Obesidad, PUERTO MONTT, Chile
[2]Universidad de Chile, Inta, Santiago, Chile

Background:
Obesity is a significant public health issue in Chile, with a prevalence of 31.4% (2017). In 2022, metabolic bariatric surgery was integrated into the public health system (Fonasa), improving accessibility to this treatment. However, its implementation has revealed gaps in postoperative monitoring within primary care, highlighting the need to identify patients in preoperative patients to optimize clinical follow-up.

Objectives:
To establish a descriptive baseline as reference for a future postoperative surveillance system, a multivariate analysis and machine learning approach will be applied to identify patterns in pre-metabolic bariatric surgery patients.

Methods:
A cross-sectional observational study was conducted with 116 preoperative patients from southern Chile (98 women, 18 men, mean age 37.97 ± 9.84 years, mean BMI 40.22 ± 4.29 kg/m²). Principal component analysis (PCA) was applied to reduce the dimensionality of clinical, metabolic, and demographic variables. Subsequently, hierarchical clustering with Gower's distance was used to identify homogeneous subgroups. The clusters were validated using internal quality indices and bootstrap procedures that assessed the stability of the assignments. Support vector machines (SVM) and multilayer perceptron (MLP) were used to estimate the ability of the algorithms to reproduce the structure found and thus provide evidence of the consistency of the classification.

Results:
The clustering approach identified subgroups with similar metabolic profiles, which served as a baseline for patient classification. The analysis showed associations between preoperative metabolic characteristics and subgroup allocation. These results provided a framework for stratifying patients before surgery and a basis for designing postoperative surveillance strategies.

Conclusions:
The use of multivariate and machine learning methods in the classification of preoperative patients generated a baseline for a postoperative surveillance system. The combination of hierarchical clustering, SVM, and MLP allowed the identification of metabolic subgroups, the evaluation of cluster stability, and the estimation of the capacity to reproduce the observed structure. These results supported the formulation of monitoring strategies and the design of subsequent study phases.

Disclosure:
No significant relationships.

General objective

To identify **preoperative patterns** in bariatric surgery patients through

Multivariate analysis and **Machine learning** using demographic, clinical, and paraclinical data.

Establish groups that serve as a **baseline**

for the subsequent development of a postoperative monitoring system in primary care.

Variable

Demógraficas

Edad

Sexo

Metábolicas

Hemoglobin
Hematocrit
Mean Corpuscular Volume
Serum Ferritin
Transferrin
Iron Saturation,
Total Iron Binding Capacity (TIBC)
Total Iron
Plasma Iron
UIBC

HDL cholesterol
LDL cholesterol
Total cholesterol
Triglycerides
GOT (AST)
GPT (ALT)
Alkaline phosphatase
GGT
Serum calcium
Serum phosphorus

TSH
Glicemia en ayunas
Insulina basal
Indice HOMA-IR
HbA1c
Albúmina sérica
Proteínas totales
Vitamina B12
Vitamina D
Ácido úrico

Clínicas

Antropométricas

Peso
Talla
IMC

Cirugía previa

Consumo de Alcohol
Consumo de tabaco
Consumo de drogas

Alergia a penicilina

Comorbilidades

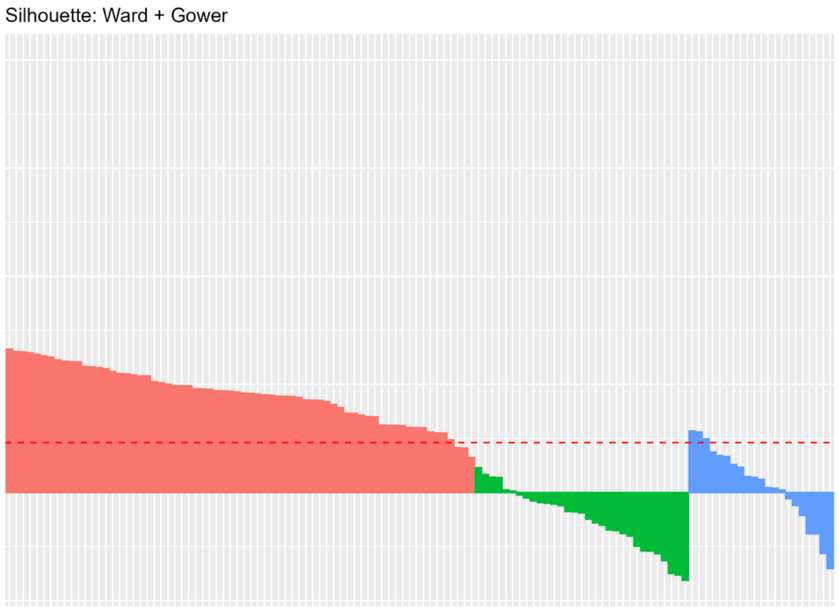
Diabetes mellitus
Prediabetes
Hipertensión arterial
Dislipidemia
Hígado graso
Síndrome metabólico
Hipotiroidismo
Reflujo gastroesofágico
Resistencia a la insulina
Hernia lumbar
Esofagitis endoscópica
Gastropatía endoscópica
Incompetencia cardial
Hipertrofia ventricular
Colecistolitiasis

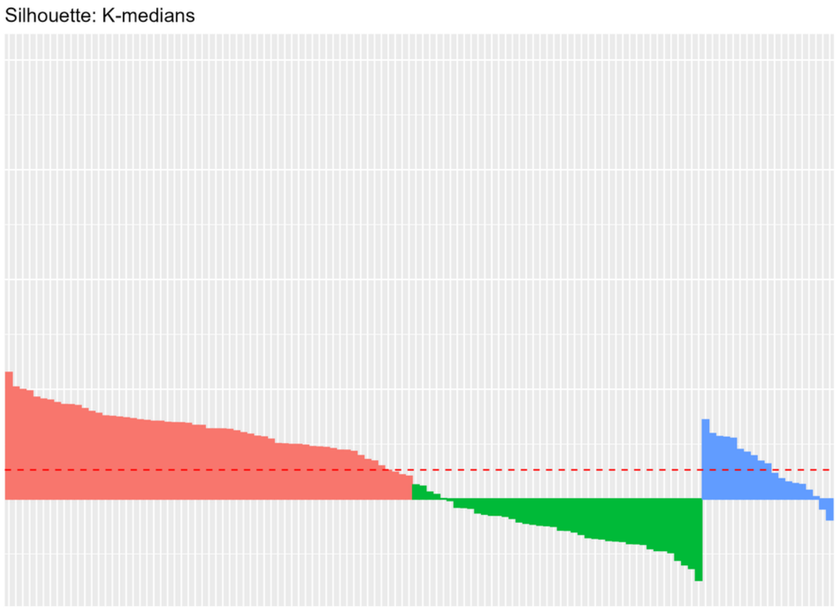
Uso de Médicamentos

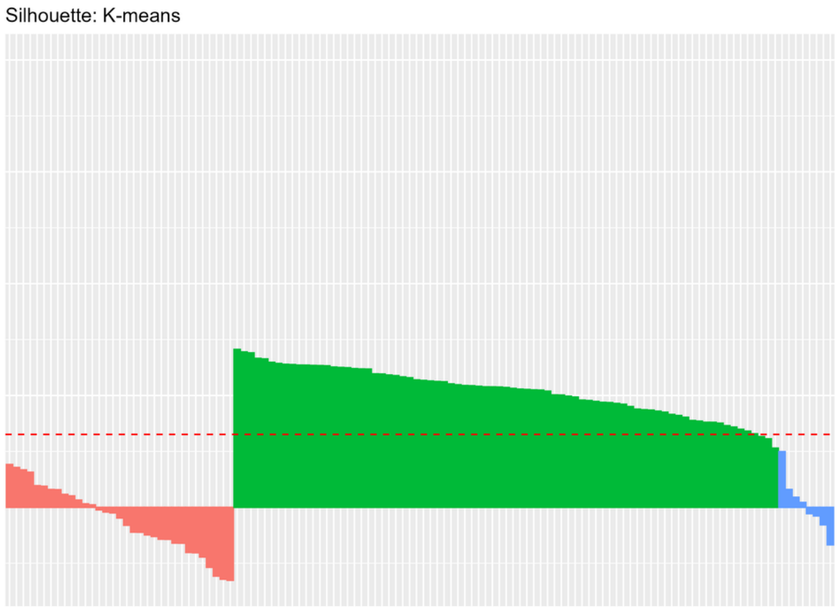
Levotiroxina
Losartán, enalapril, amlodipino
Estatinas
Metformina
Inhibidores del cotransportador de sodio-glucosa tipo 2

Tipo de círugia

Silhoutte Index







The three panels show silhouette index plots for Ward + Gower, K-means, and K-medians.
The objective is to observe the internal consistency of each group and incorrectly assigned values.

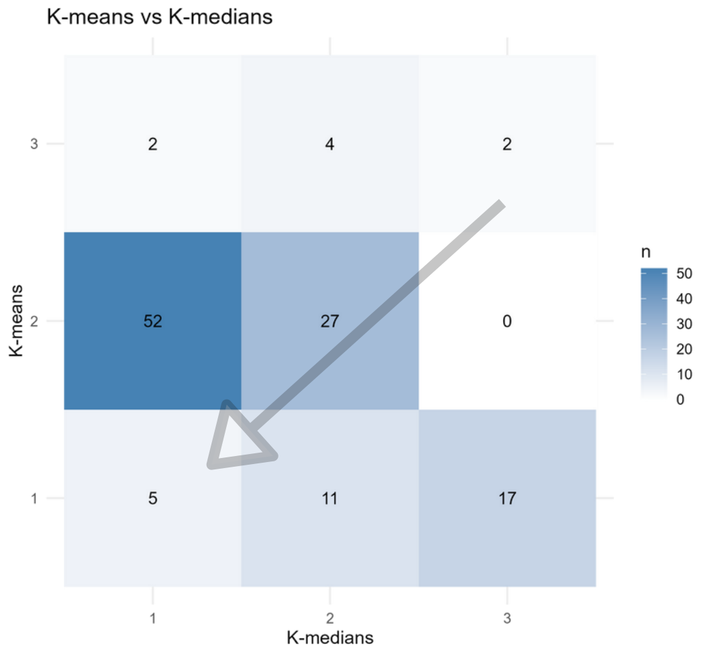
- Ward + Gower and K-medians: **many negative values, weak cluster separation.**
- K-means: mostly positive values, stronger and clearer clustering.

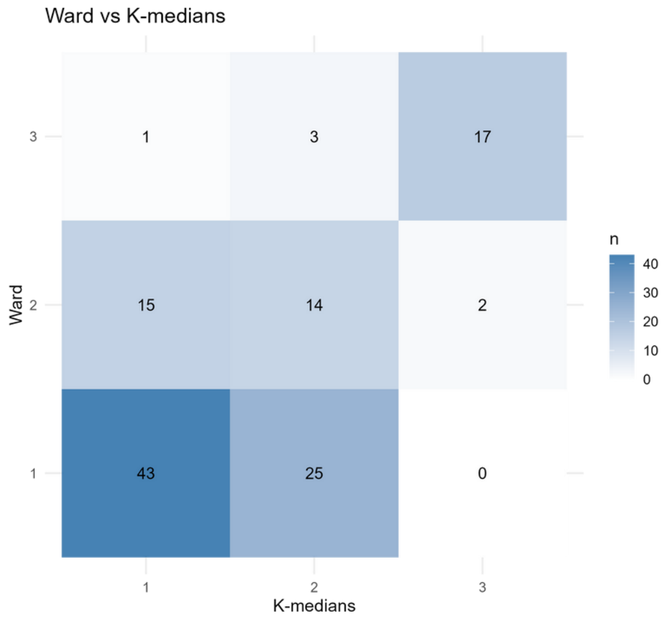
The silhouette index measures how well each point fits in its cluster:

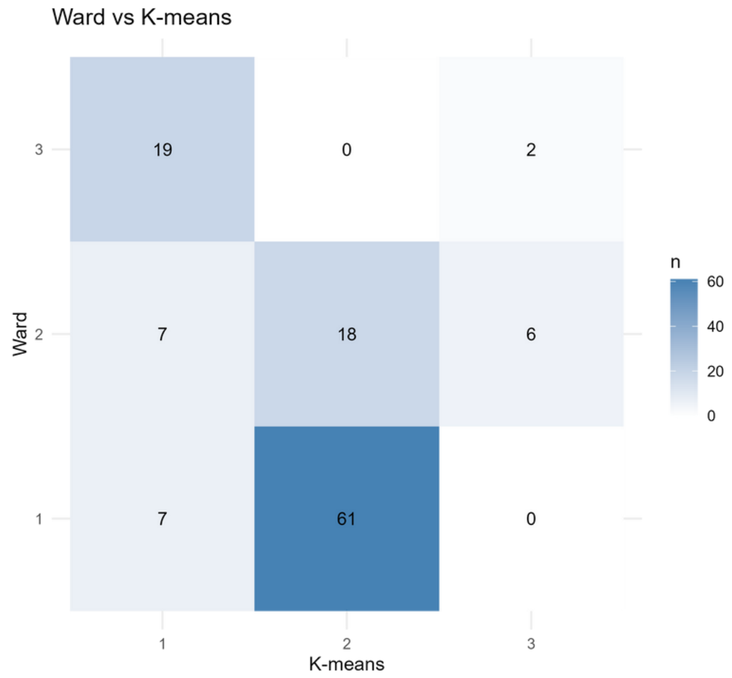
- values near +1 are well clustered
- near 0 are ambiguous
- below 0 suggest misclassification.

Confusions Matrix

Comparing the results in the matrices identified similarities and discrepancies in the classifications.







They agree in some groups but differ in borderline observations.

Adjusted Rand Index (ARI)

A consensus of clusters was built:
1.Ward+Gower
2.K-means
3.K-medians.

Two observations receive a value of 1 if they fall into the same group and zero otherwise.

↓

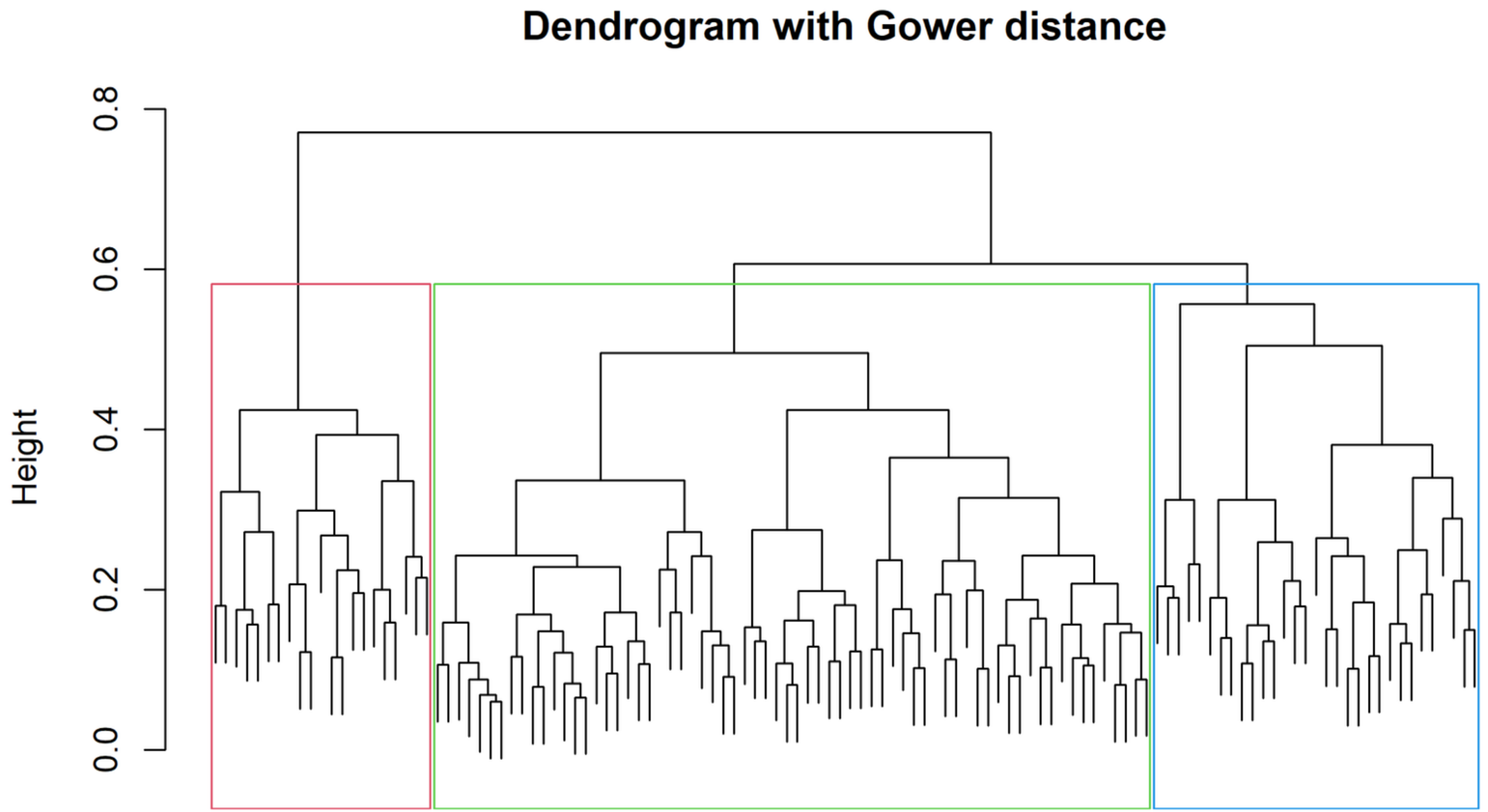
Average.

El ARI (Adjusted Rand Index) measures similarity between partitions by random adjustment (1 = perfecto (0 = azar)

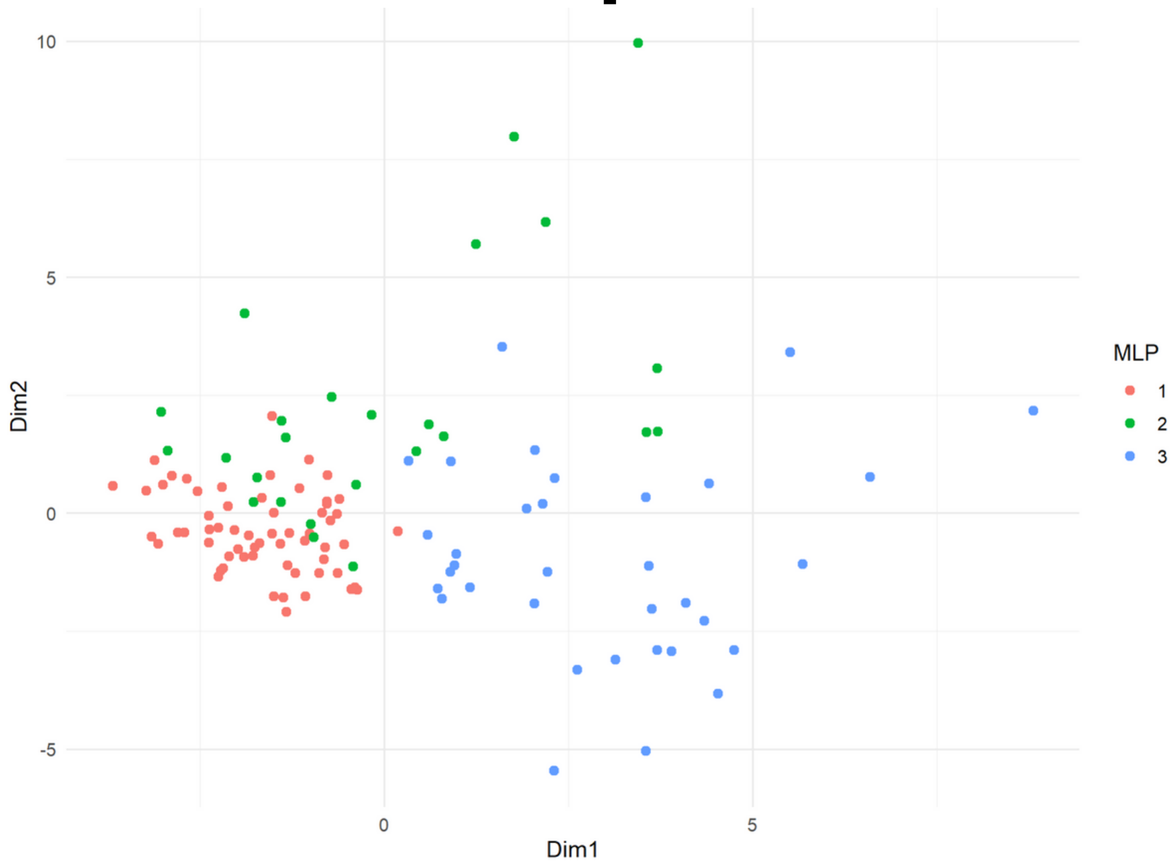
Resultados (ARI vs consenso):
Ward+Gower → 0.739 (alta similitud)
K-means → 0.627 (moderada)
K-medians → 0.197 (baja)

Hierarchical classification (Ward Linkage)

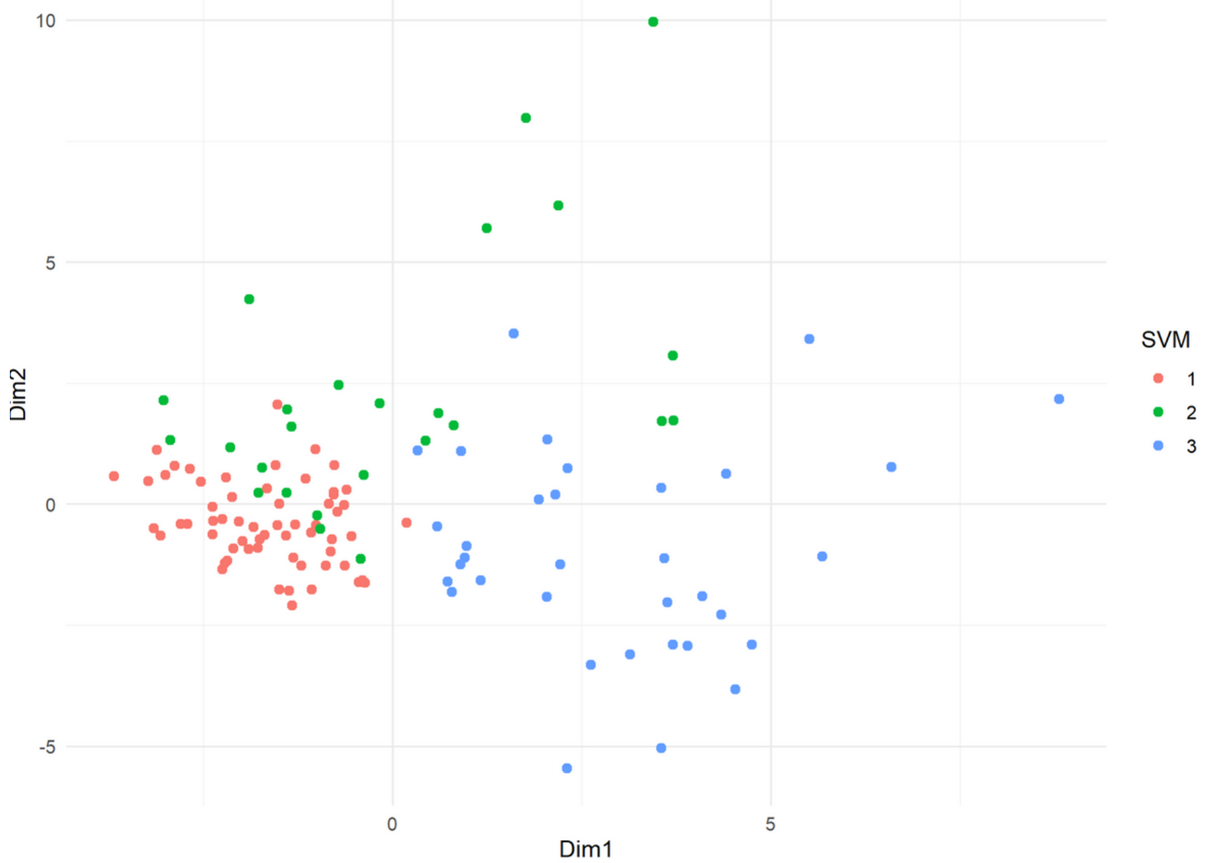
averiguar que es el higjt



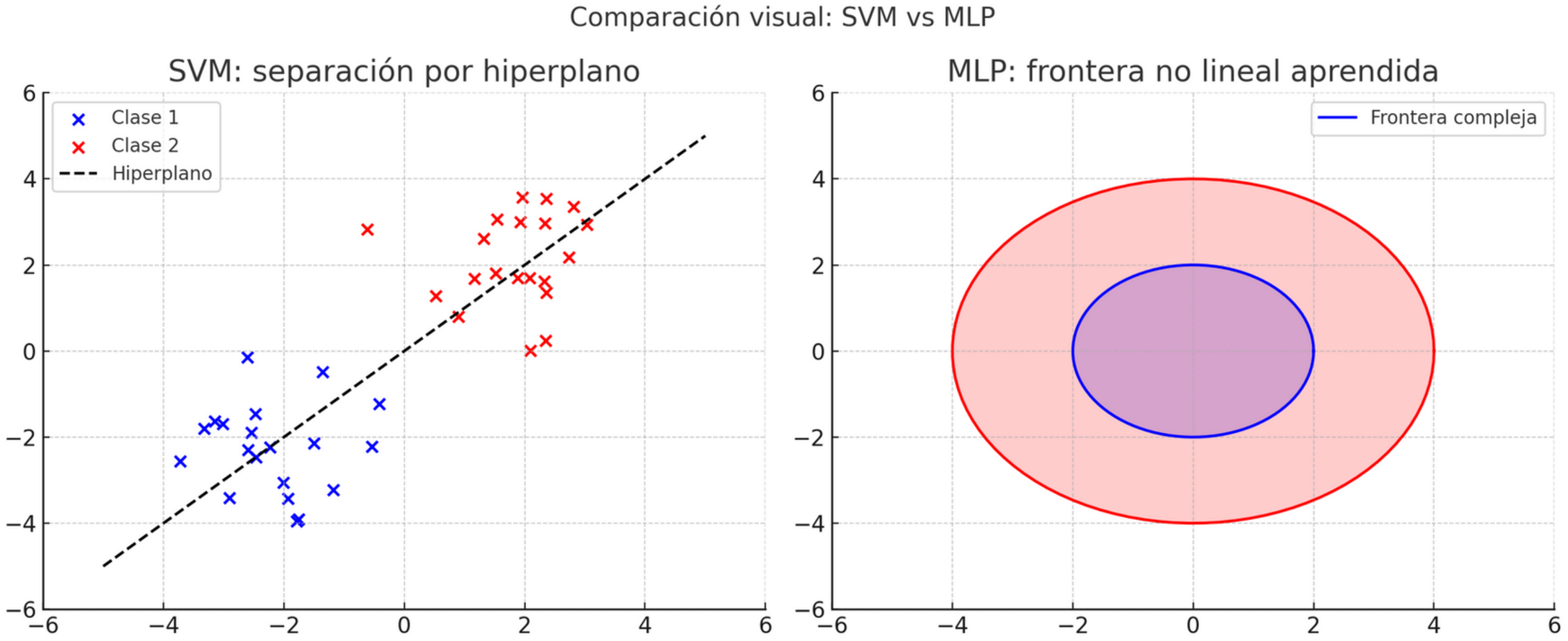
Multilayer Perceptron

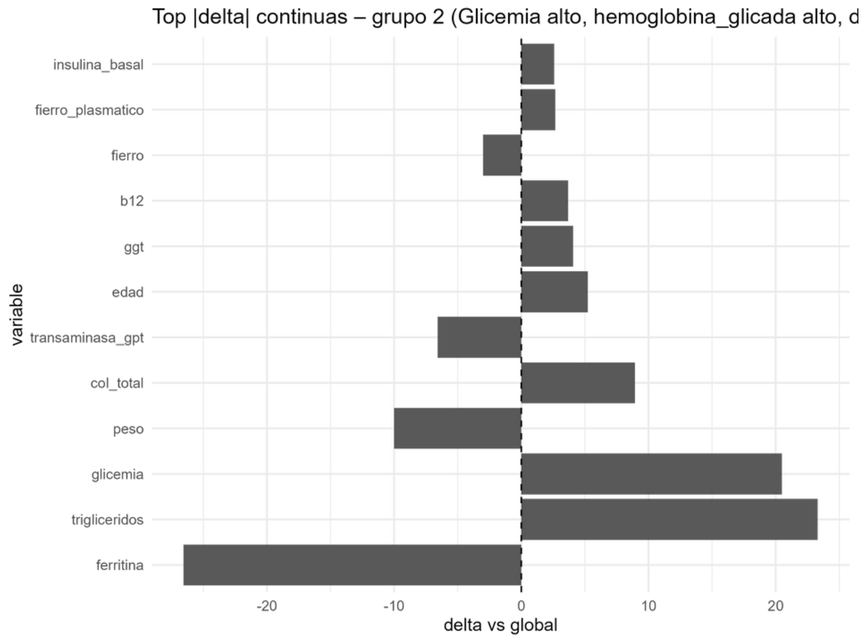


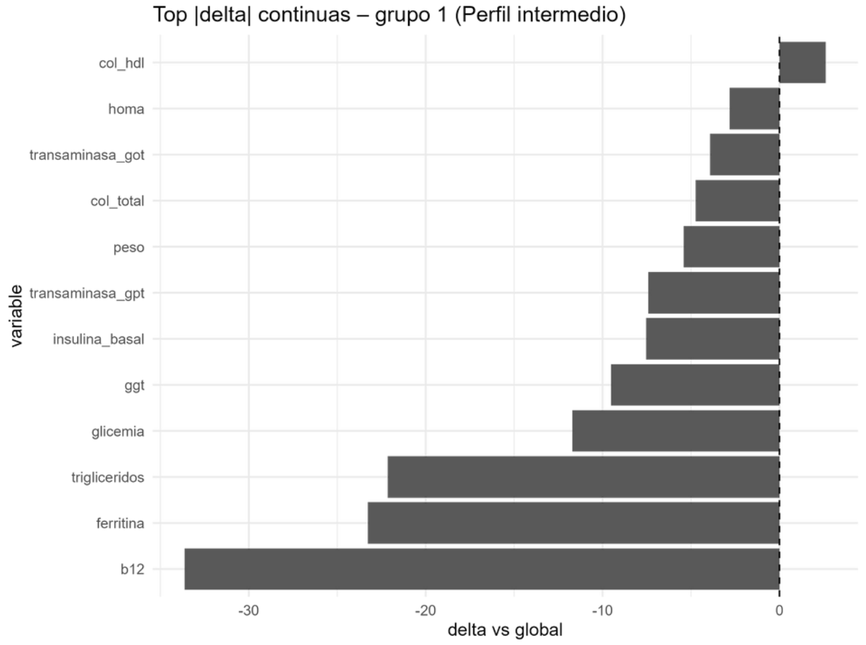
Support Vector Machine

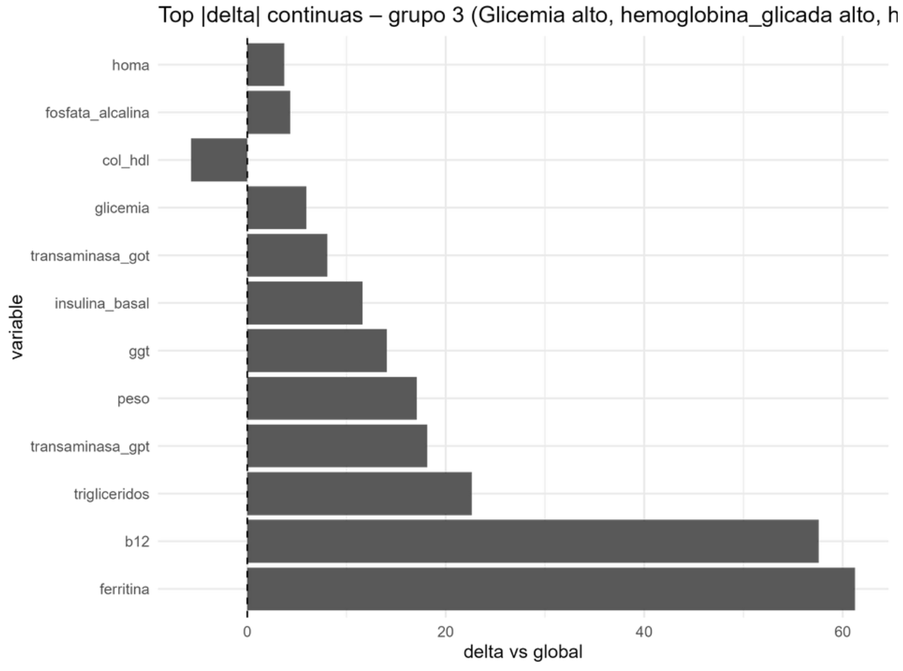


Support Vector Machine and Multilayer Perceptron were trained with principal components to replicate consensus and classify new observations generating three new groups.


Comparación visual: SVM vs MLP
SVM: separación por hiperplano
Clase 1
Clase 2
Hiperplano
MLP: frontera no lineal aprendida
Frontera compleja


Top |delta| continuas – grupo 2 (Glicemia alto, hemoglobina_glicada alto, d
insulina_basal
fierro_plasmatico
fierro
b12
ggt
edad
transaminasa_gpt
col_total
peso
glicemia
trigliceridos
ferritina
variable
-20
-10
0
10
20
delta vs global


Top |delta| continuas – grupo 1 (Perfil intermedio)
col_hdl
homa
transaminasa_got
col_total
peso
transaminasa_gpt
insulina_basal
ggt
glicemia
trigliceridos
ferritina
b12
variable
-30
-20
-10
0
delta vs global


Top |delta| continuas – grupo 3 (Glicemia alto, hemoglobina_glicada alto, h
homa
fosfata_alcalina
col_hdl
glicemia
transaminasa_got
insulina_basal
ggt
peso
transaminasa_gpt
trigliceridos
b12
ferritina
variable
0
20
40
60
delta vs global