

## INTRODUCCIÓN

- Actualmente el Síndrome Metabólico (SM) es una entidad bien reconocida y de alta prevalencia a nivel mundial.
- Se encuentra relacionado a enfermedades crónicas como la diabetes, hígado graso no alcohólico y la presencia retinopatía.
- Su diagnostico es caro y/o invasivo.
- El detectar trastornos metabólicos previos a las complicaciones del SM ayudará al tratamiento oportuno.
- Escogimos el algoritmo de árbol de decisión por ser similar a la toma de decisiones por un medico al momento de diagnosticar una enfermedad.

## OBJETIVOS

- Clasificar las complicaciones del síndrome metabólico mediante el uso del árbol de decisión con variables bioquímicas y metabolómicas.
- Uso de árbol de decisión para clasificación por su facilidad de interpretación.
- Mejorar el algoritmo de decisión par automatizar el diagnostico de las complicaciones del síndrome metabólico.
- Seleccionar las características mas importantes mediante el método de paso hacia delante, Neighborhood components analysis, y análisis de factores, reduciendo así las dimensiones.
- Generar meta características para incrementar la eficiencia del árbol de decisión para clasificar las complicaciones del SM.

## MATERIAL Y MÉTODOS

Base de datos	Fuente	E/C	Variables
Diabetes	UANL	40/40	57
NAFLD	UNAM / NIH	31/106	82
Retinopatía	UNAM / NIH	39/19	82

Los datos fueron analizados mediante R y Matlab.

```

Generalized Cost Sensitive Look Ahead Decision Tree (GCSLADT)(D,d)
% D: conjunto de datos, d: tamaño de profundidad
Input : atributos valuados en el conjunto de datos D
Output: Un GCSLADT
if D es "puro" u otro criterio de paro se cumple then
    termina
end
forall Atributos ∈ D do
    Computar el criterio de información teórico si particionamos en a
end
amejor = El mejor subconjunto de atributos de tamaño d, acorde a los criterios de
información teórico computados antes
Árbol = Crea una rama de decisión que prueba amejor en la raíz
DI = Inducción de los sub-datos de D basados en amejor
forall DI do
    Árbolv = (GCSLADT)(DI, d)
    Adjunta el árbolv a la rama correspondiente del árbol
end
return Árbol
  
```

## CRITERIO DE INFORMACIÓN

**Información teórico:** Encontrar un atributo con un valor de limite que maximiza la reducción de entropía para realizar la partición de datos.

$$(k^*, \tau^*) = \operatorname{argmax}_{k, \tau} G(k, \tau)$$

Donde  $G(k, \tau) = E_{\text{antes}} - E_{\text{despus}}$ .  $E$  denota la entropía, que es una medida de incertidumbre en los datos.

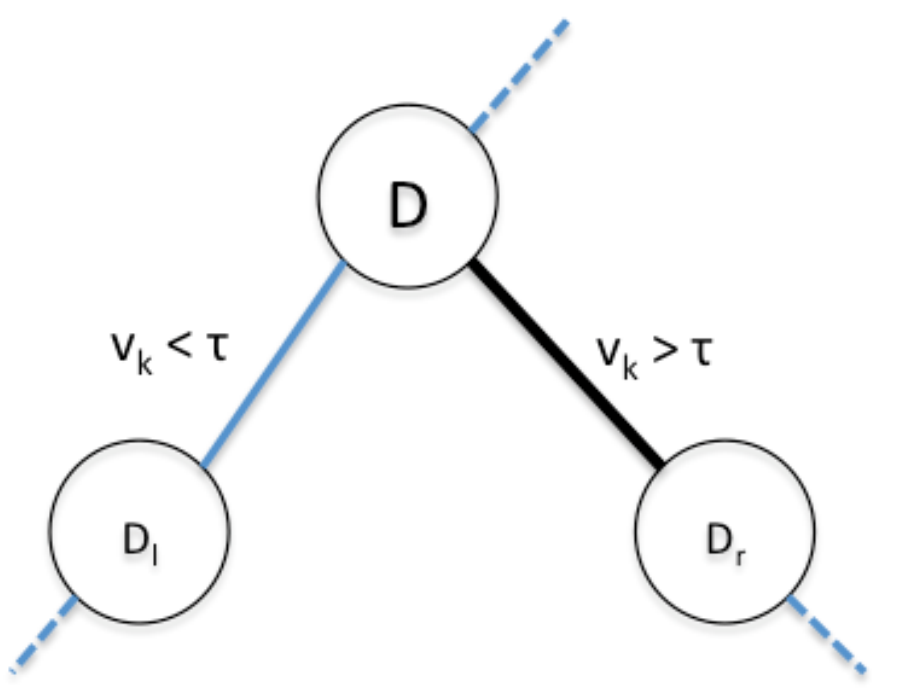
$E_{\text{antes}}$  y  $E_{\text{despus}}$ , la entropía antes y después de la partición.  $E_{\text{antes}} = E(D) = \sum_{i=1}^m f(w_i, D) p(c_i, D) \log_2(p(c_i, D))$

$$E_{\text{despus}} = \frac{|D_l(k, \tau)|}{|D|} E(D_l(k, \tau)) + \frac{|D_r(k, \tau)|}{|D|} E(D_r(k, \tau))$$

Árbol de decisión sensible a costos

$$\frac{f(w_i, D)}{\sum_{i=1, c \neq i}^m |D(c_i)|} =$$

$|D(c_i)|$  = # de observaciones en D pertenecientes a clase  $c_i$ .  $f(w_i, D) = 1 \forall i$  para árboles de decisión no sensibles a costo



## META-FEATURES

Las meta-características resultan de aplicar, una operación algebraicas o booleanas entre las variables de la base de datos.

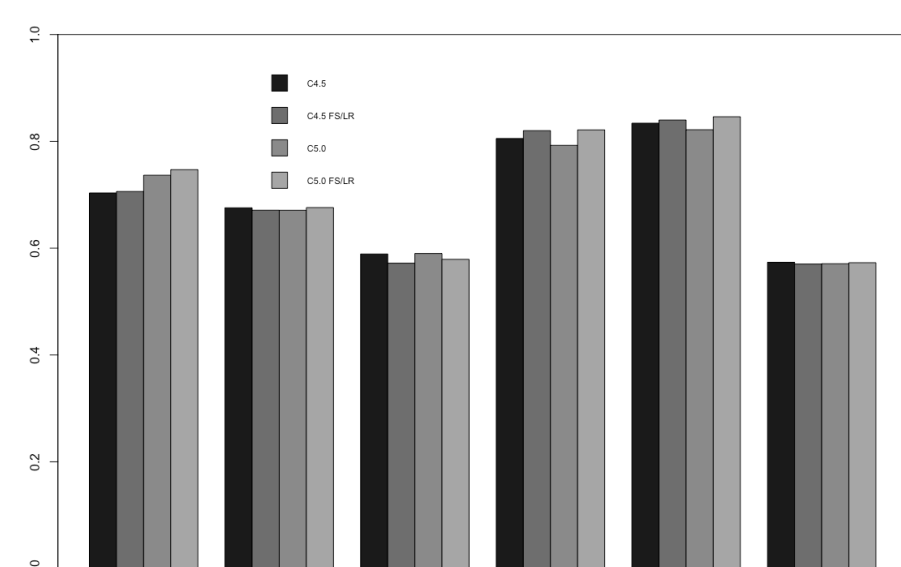
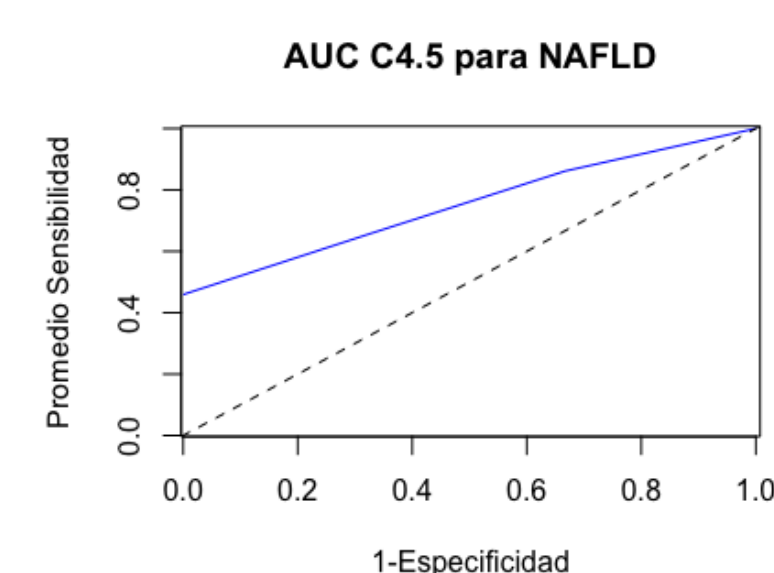
	Bioquímico	Bioquímico + Metadata
Exactitud	83%	88.3%
# Hojas	7	13
T de árbol	7	13

## SELECCIÓN DE CARACTERÍSTICAS

Método de Selección	Forward selection + logistic regression	NCA + DT	Factor analysis + DT
Exactitud	90%	81%	87.5%
# Hojas	7	7	7
T de árbol	13	13	13

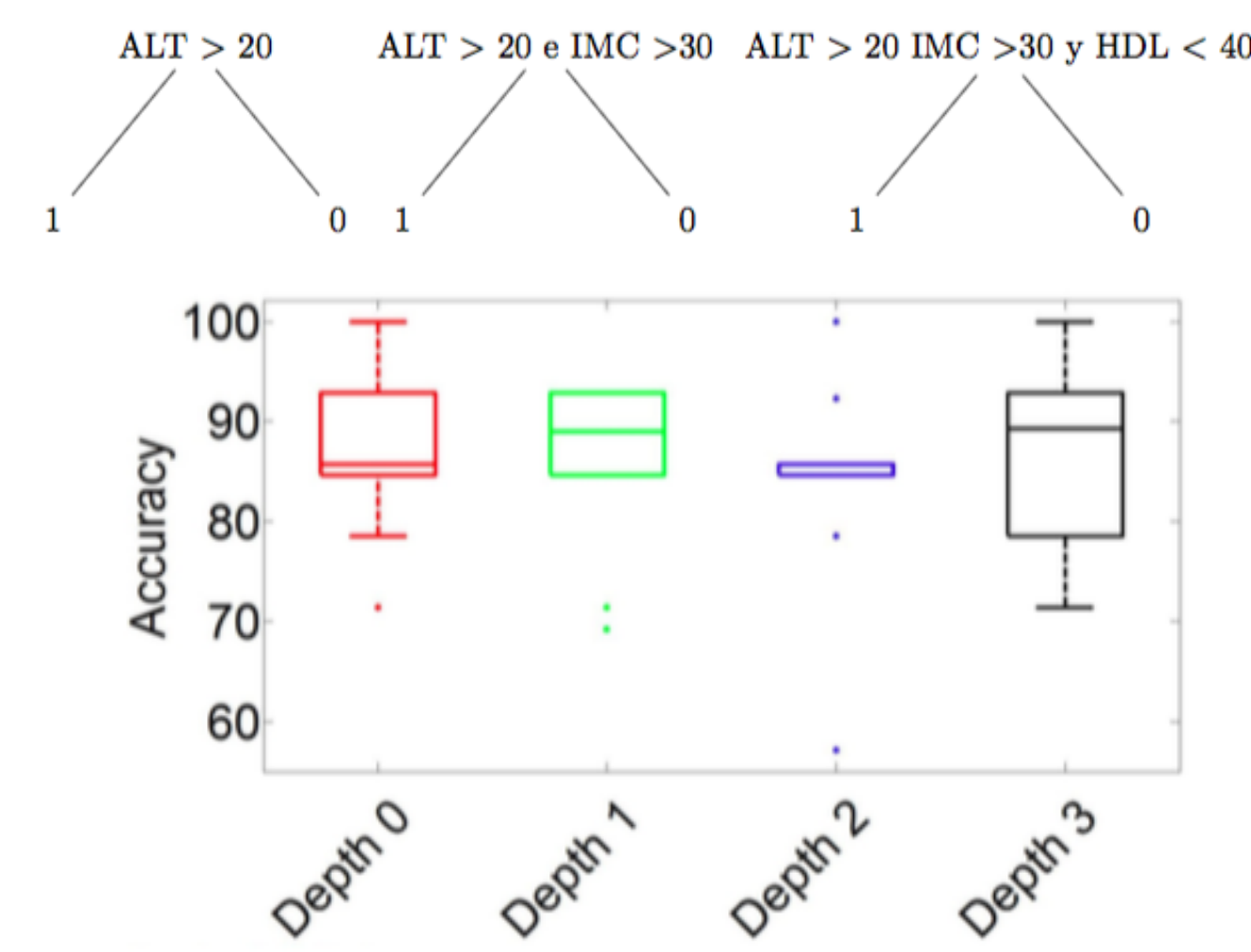
## RESULTADOS

C4.5	Exactitud	AUC
Diabetes	91.2%	90%
NAFLD	97.6%	95.9%
Retinopatía	91.6%	94.6%

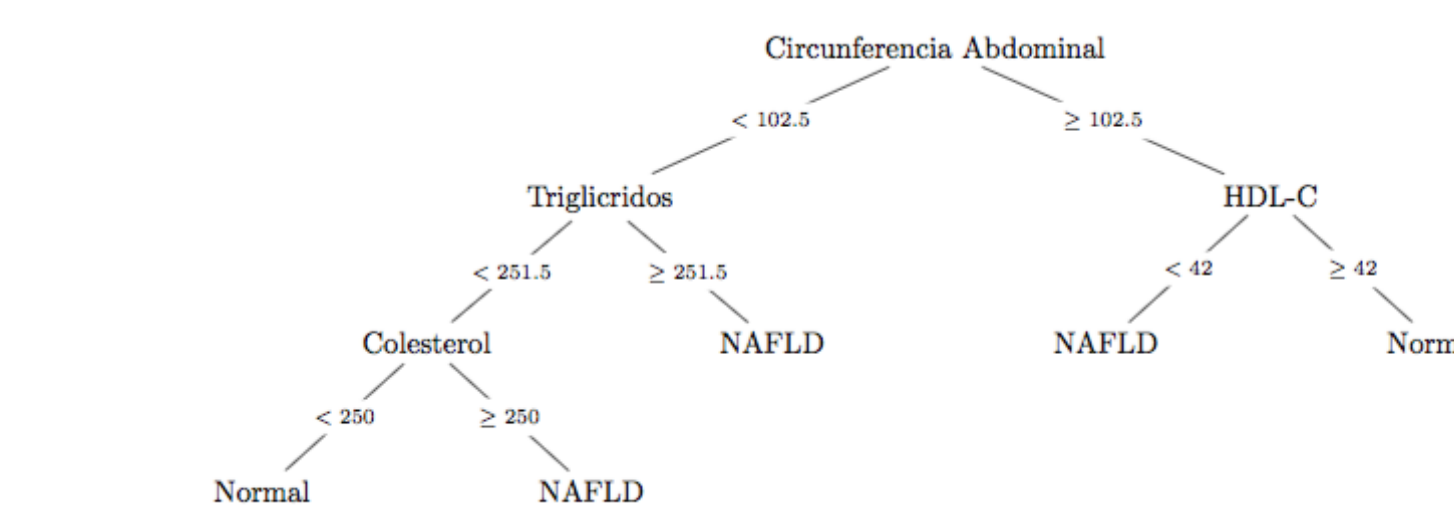


## LOOK AHEAD DECISION TREE

En este algoritmo se toma en cuenta los nodos adelante del nodo actual para tomar la decisión de clasificación.



## COST SENSITIVE DECISION TREE



E(NAFLD)/ A(Normal)	E(Normal)/ A(NAFLD)	Exactitud
0.25	0.75	91.42
0.75	0.25	91.42
0.33	0.67	92.22
0.67	0.33	91.72
0.5	0.5	91.24

## CARACTERÍSTICAS Y RESULTADOS

Bioquímicas	Inflamatorias	Metabolómicas
Método	PCR	Actividades
Al.Úrico	GM-CSF	Arginina
Colesterol total	IFNγ	Glicina
Tríglicéridos	IL10	Alanina
HDL-C	IL2	Leucina
LDL-C	IL4	Succinil carnitina
ALT	IL8	Citulina
AST	TNFα	Metionina
Glucosa		Fenilalanina
Urea		CSDC_C4DC
Creatinina		Oxalato
		Prolina
		Valina

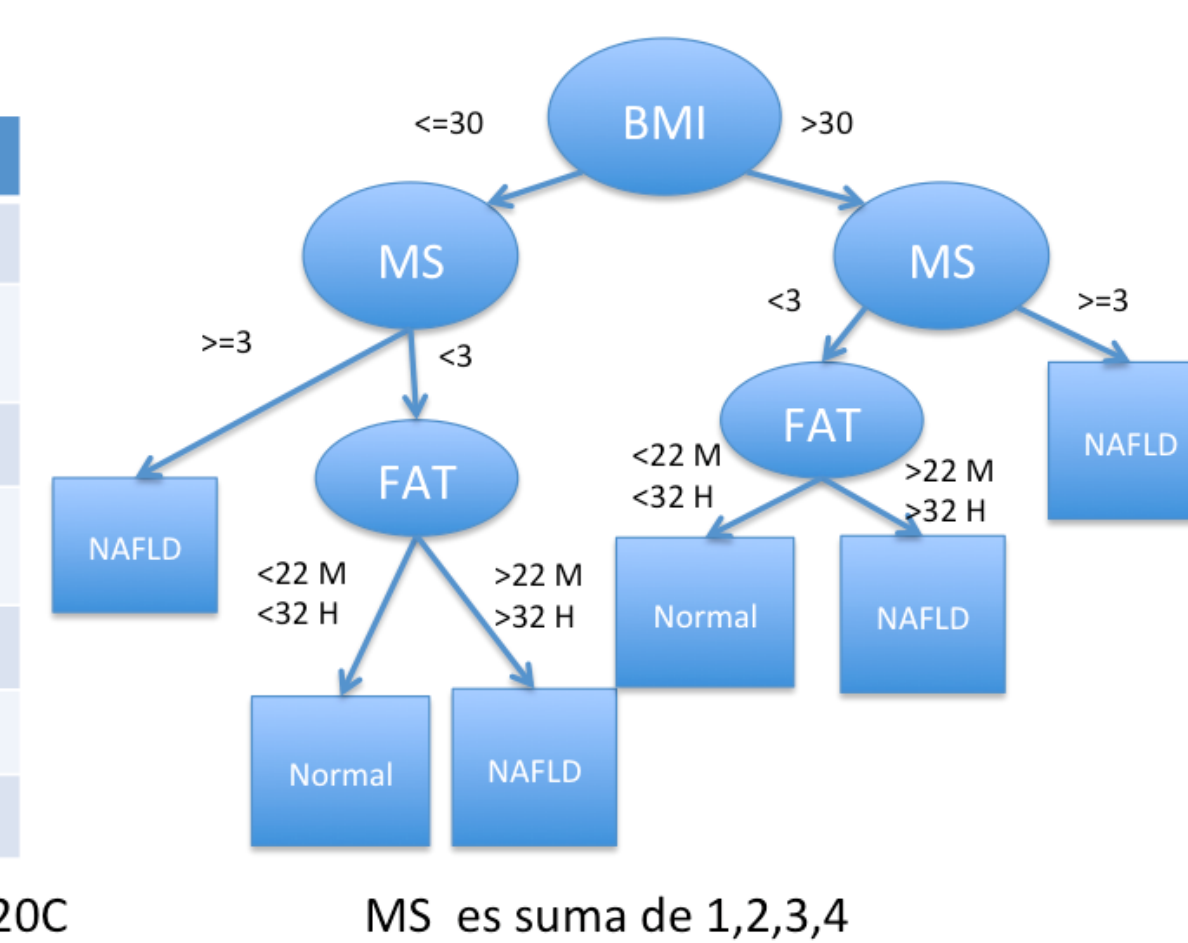
C4.5 / NAFLD	Exactitud	ROC
FS/LR	90%	93%
Bioquímico + Metadata	88.3%	81%
DT / Análisis de factores	87.5%	89.5%
Metadata/ FS	86%	90%

## ATP III SÍNDROME METABÓLICO + FAT + BMI

**Criterio de Síndrome Metabólico**

- Tríglicéridos > 150 mg/dl
- HDL -c: Hombres < 40 mg/dl, Mujeres < 50 mg/dl
- Glucosa < 100 mg/dl
- Circunferencia abdominal: Hombres > 102 cm, Mujeres > 88 cm
- Presión arterial sistólica: > 130 y diastólica > 85
- FAT %: Hombres > 32, Mujeres > 22

Am J Cardiol. 2002 Mar 7;89(5A):8C-20C



	Sens	Esp	VPP	VPN	Exactitud
Bioquímico	75%	89%	64.5%	91.5%	83%
Bioquímico + Metadata	80%	90%	64%	95%	88.3%
FS / LR	82%	92%	74%	95%	90%
ATPIII SM	23%	78%	70%	31%	40.14%

## CONCLUSION

- GCSLADT supera otras variantes de árbol de decisión para el diagnóstico de complicaciones del síndrome metabólico.
- La incorporación de meta características ayuda a dar mayor exactitud del árbol de decisión para clasificar las complicaciones del síndrome metabólico.
- La función de costos clase sensible también incrementa la exactitud del árbol de decisión para clasificar el NAFLD porque los datos no son balanceados.
- GCSLADT puede demostró tener una mejor exactitud que el utilizar un árbol de decisión creado por el experto para clasificar NAFLD.
- El algoritmo de árbol de decisión pudo clasificar adecuadamente al paciente con diabetes, NAFLD y retinopatía con las acilcarnitinas de cadena larga.

## REFERENCIAS

- Thierry Poynard, Vlad Ratziu, Sylvie Naveau, Dominique Thabut, Frederic Charlotte, Djamila Messous, Dominique Capron, Annie Abella, Julien Massard, Yen Ngo, et al. The diagnostic value of biomarkers (steatitest) for the prediction of liver steatosis. *Comparative hepatology*, 4(1):10, 2005.
- H Haller. Epidermiology and associated risk factors of hyperlipoproteinemia. *Zeitschrift fur die gesamte innere Medizin und ihre Grenzgebiete*, 32(8):124-128, 1977.
- Kurt George Matthew Mayer Alberti and PZ ft Zimmet. Definition, diagnosis and classification of diabetes mellitus and its complications. part 1: diagnosis and classification of diabetes mellitus. provisional report of a who consultation. *Diabetic medicine*, 15(7):539-553, 1998.