

BIO MED 2024

LES JOURNÉES POUR L'AVENIR DE LA BIOLOGIE MÉDICALE

JEUDI 23 &
VENDREDI 24
MAI 2024

Utilisation de l'IA dans la prédiction de la PTH et de l'IPBGLR chez le patient dialysé

Yohann Missiak, Data Analyst, Biologiste Médical
Matthieu Bernier, Project Manager, Biologiste Médical

Groupe INOVIE CBM

Partenariat

- **Clinique néphrologie St EXUPERY**

- **11 implantations d'hémodialyses**

**D.GUINAULT, D.MILONGO, A.DELARCHE,
O.MILIOTO, P.DAHAN.**

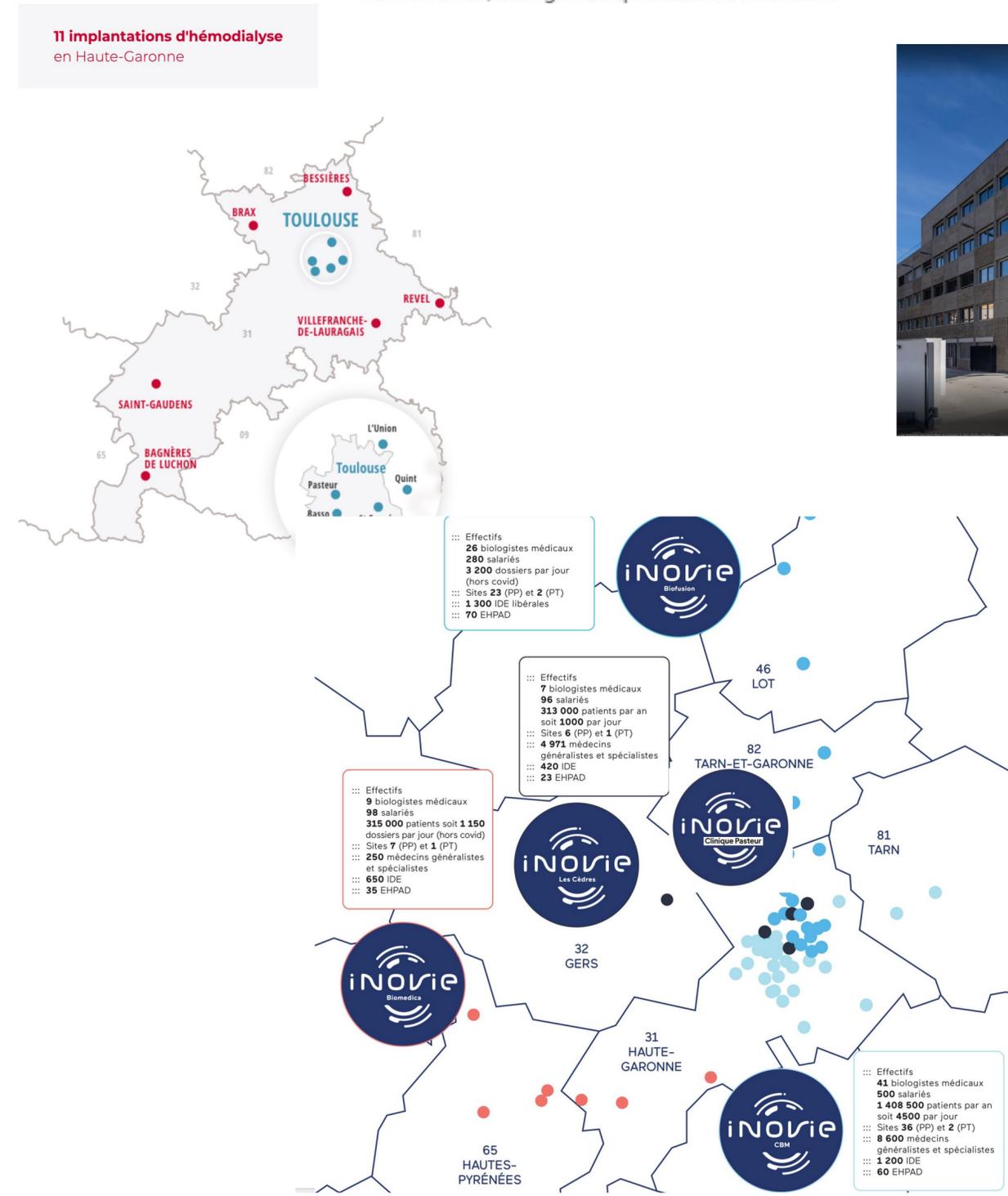
- **LBM Inovie CBM et groupe Inovie région toulousaine**

- **8 plateaux techniques**

- **94 sites**

- **11 établissements de soins**

**Y.MISSIAK, M.BERNIER, A.LECOUR,
C.ROSSIGNOL, T.JAUDON.**



Introduction

Contexte

- Suivi long terme des patients en dialyse
- Etude conceptuelle prédire PTH élevée à partir autres paramètres bilan biologique
- A terme, pipeline pour prédire différentes hormones + **conformité IA Act, EU Horizon**

Prise en charge

- Actuellement bilan simple tous les mois (pas PTH) et bilan complet tous les 3 mois (avec PTH + réticulocytes) +++
- Identification précoce d'une PTH élevée
- ->>> (Intérêt clinique majeur)

Conformité Légale de l'Etude

- **Ok Etude clinique + consentement patients + CPP + CHU IRB Montpellier**

IRB ID : 202201169 RB Accreditation number: 198711

Statistique classiques

- 136 patients dialysés avec valeur numérique PTH
- Variables démographiques (age sexe) et biologique (valeur NFS, PTH, Biochimie) + réticulocytes

Etape préparation des données

- Anonymisation +++
- Transformation variable numérique PTH en variable de classe qualitative
D'après recommandation KDIGO PTH 9 fois supérieure à valeur de référence traditionnelle
Si PTH > 331.2 pg/ml I => "PTH élevée" (Prédiction de classe)

Variables	PTH	PTH Elevée
Sodium 140, Créatinine,...	500	Oui
Sodium 132, Créatinine,...	200	Non
Sodium 142, Créatinine,...	100	Non

Matériels et Méthodes

Étapes :

- Suppression des valeurs manquantes ou inutiles vs génération des données manquantes? -> **Travail uniquement sur variables complètes+++**
- Transformation des variables catégoriques en variables numériques (0,1).++

Exploration des données

- **Statistiques classiques +++**
- permet de voir les tendances
- PCA, corrélogramme

PTH élevée % : 40%

- "Non balancé" mais très bon. 40% proche de 50%

Matériels et Méthodes

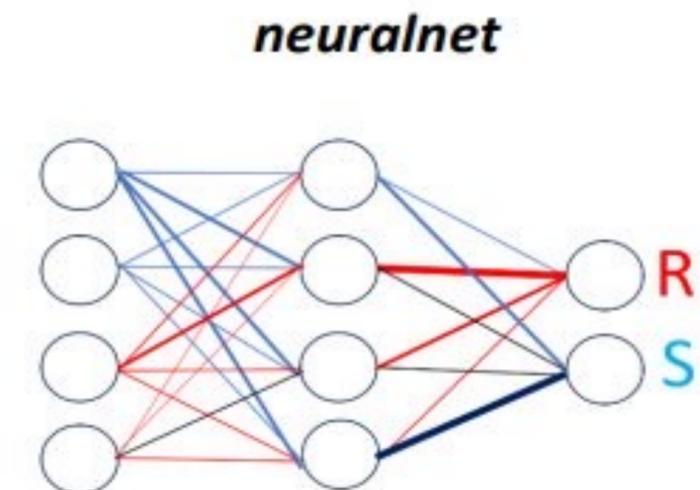
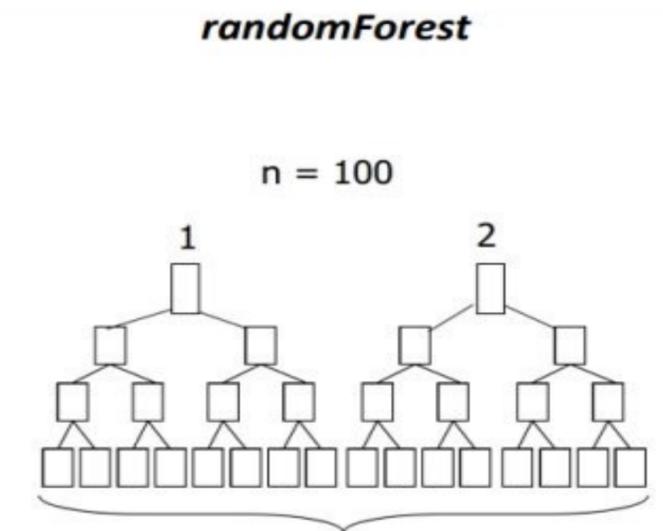
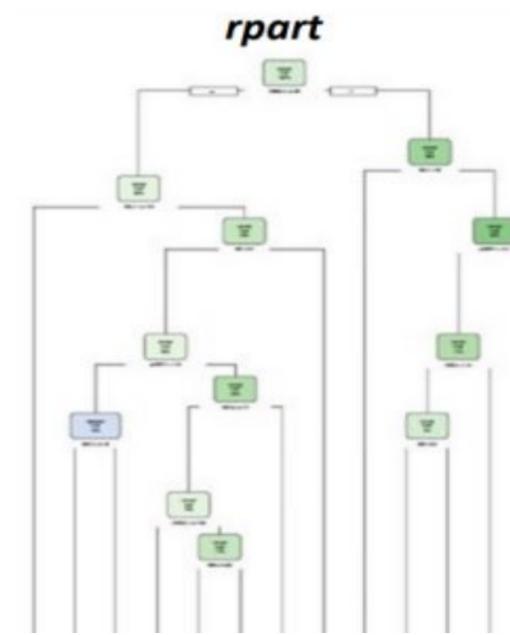
Modèles utilisés

Logiciel R

- arbres de décisionnels
- forêts randomisées
- réseaux neuronaux

Séparer les données :

- Séparer le jeu de données en deux parties :
- Une partie pour l'entraînement (par exemple, 80%)
- Une partie pour la validation (par exemple, 20%).
- -> **Evaluer la performance du modèle sur des données non utilisées lors de l'entraînement**



Entraînement

- Les modèles ont accès à la colonne "PTH Elevée" (oui/non)
Ils cherchent les liens entre la colonne "PTH Elevée" et les autres variables...

Créatinine	Sodium		PTH Elevée
20	140	<== MODELE ==>	Oui
40	138	<== MODELE ==>	Non

Validation

- On présente un nouveau jeu de donnée aux modèles (sans la colonne "PTH Elevée") et le modèle essaye de prédire la valeur (oui/non) de la colonne "PTH Elevée")

Créatinine	Sodium		Estimation PTH Elevée
23	138	MODELE ==>	Pense que oui
42	130	MODELE ==>	Pense que non

Matériels et Méthodes

Validation

- Puis on compare l'estimation du modèle à la vraie valeur réelle de "PTH Elevée"
- Extraction du taux de bonnes réponses : précision + vrais positifs, vrais négatifs,...

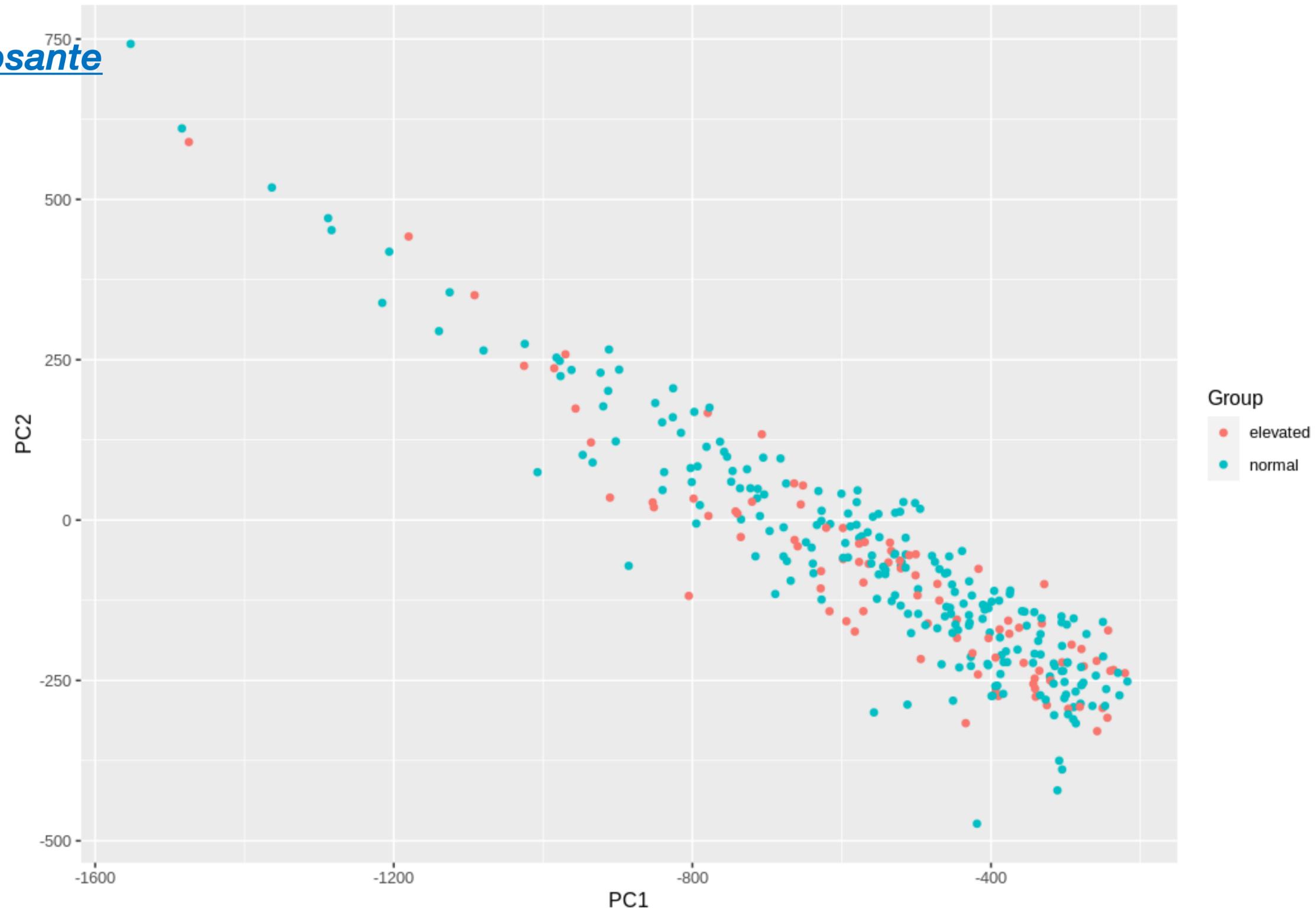
Estimation "PTH Elevée"	Vraie valeur "PTH Elevée"	Bonne réponse?
Pense que oui	Oui	Oui (VP)
Pense que non	Non	Oui (VN)
Pense que non	Oui	Non (FN)
Pense que oui	Non	Non (FP)
Pense que oui	Oui	Oui (VP)
		Total : 3/5 =précision =accuracy

	Réel OUI	Réel NON
Pense OUI	10	0
Pense NON	0	10

Résultats : Exploration des données

Analyse en Composante Principale

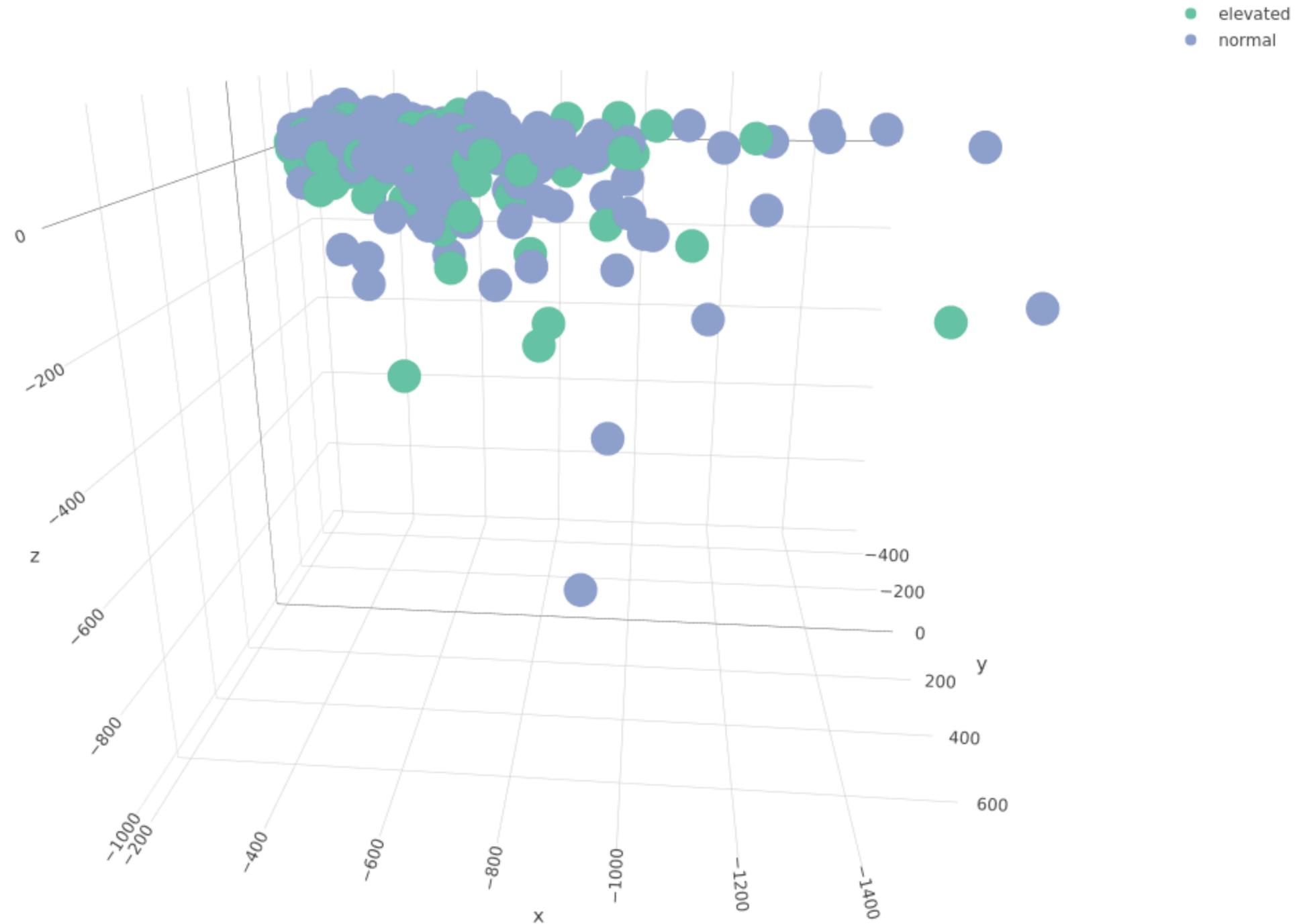
PC1
PC2



Résultats : Exploration des données

Analyse en
Composante
Principale

X=PC1
Y=PC2
Z=PC3

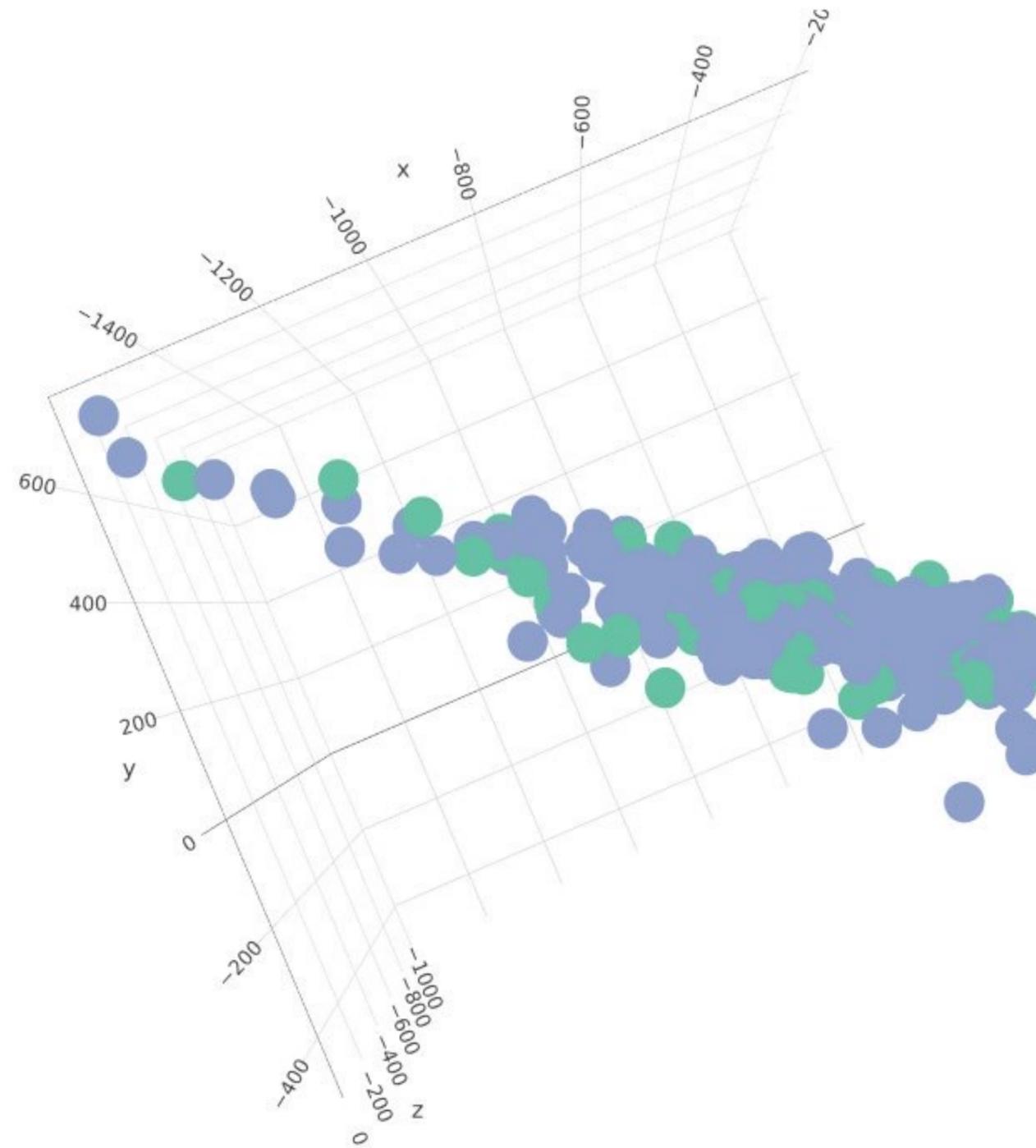


3D

Résultats : Exploration des données

Analyse en Composante Principale

X=PC1
Y=PC2
Z=PC3



● elevated
● normal

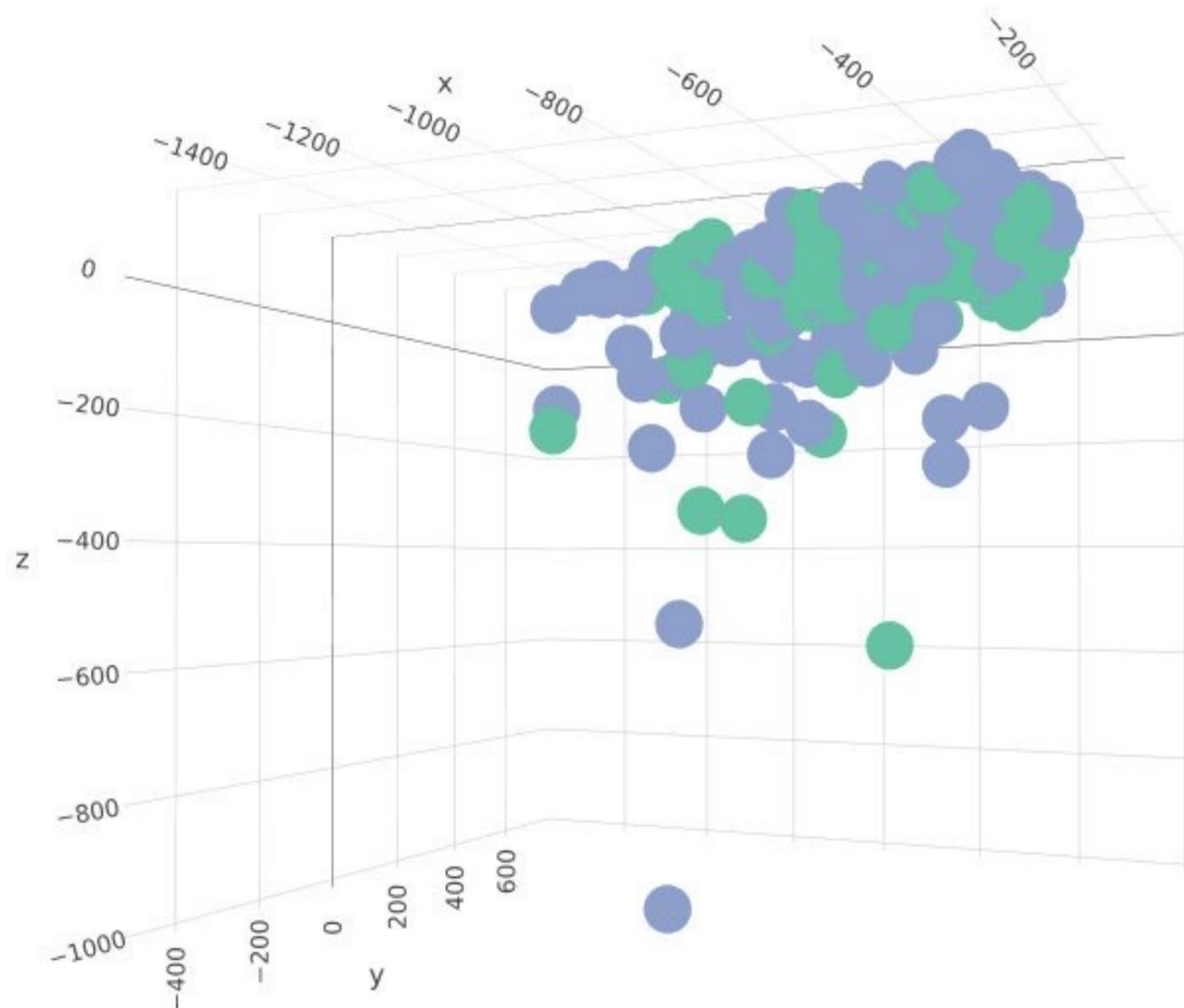
3D

Résultats : Exploration des données

Analyse en Composante Principale

X=PC1
Y=PC2
Z=PC3

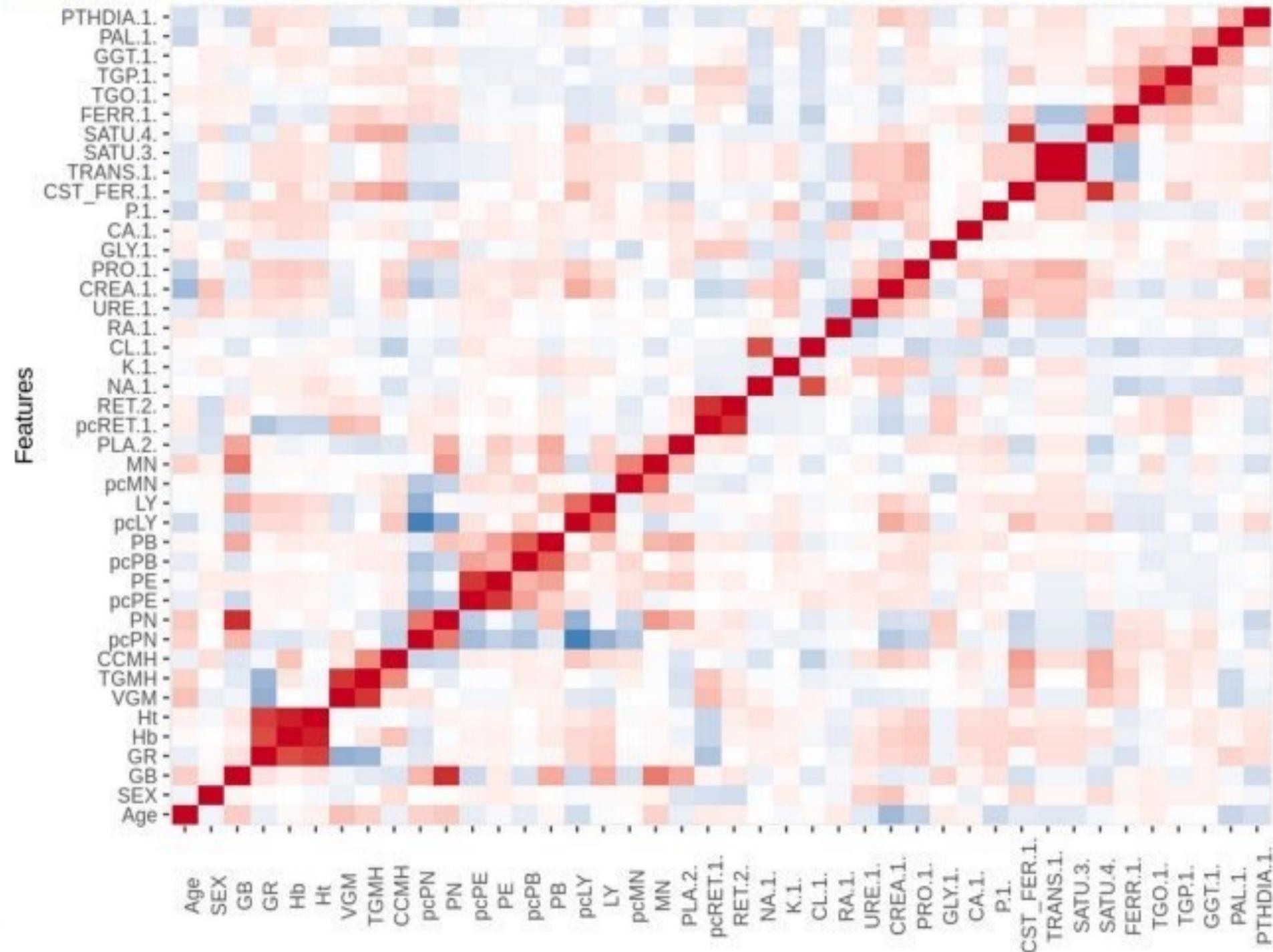
● elevated
● normal



3D

Résultats : Exploration des données

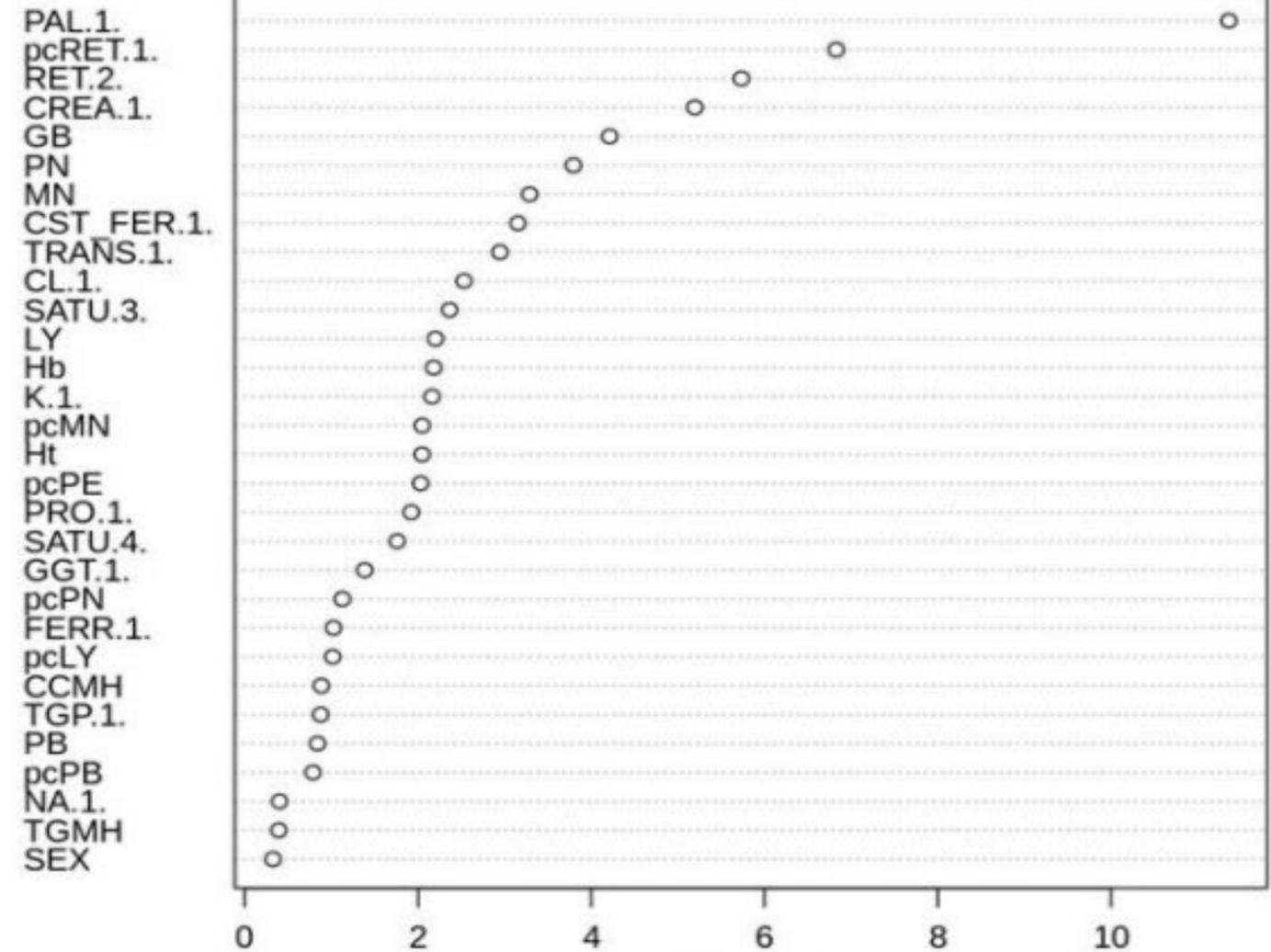
Figure 1 : Corrélogramme des paramètres de biologie et âge du patient dialysé (n=136)



Résultats : Performance

Modèle	Arbre D	Fôrets R	NN
Précision	75 %	88 %	62 %
Sensibilité	71 %	60 %	45 %
Spécificité	76 %	95 %	71 %
VPP	45 %	75 %	45 %
VPN	90 %	90 %	71 %
AUC	73 %	78 %	58 %

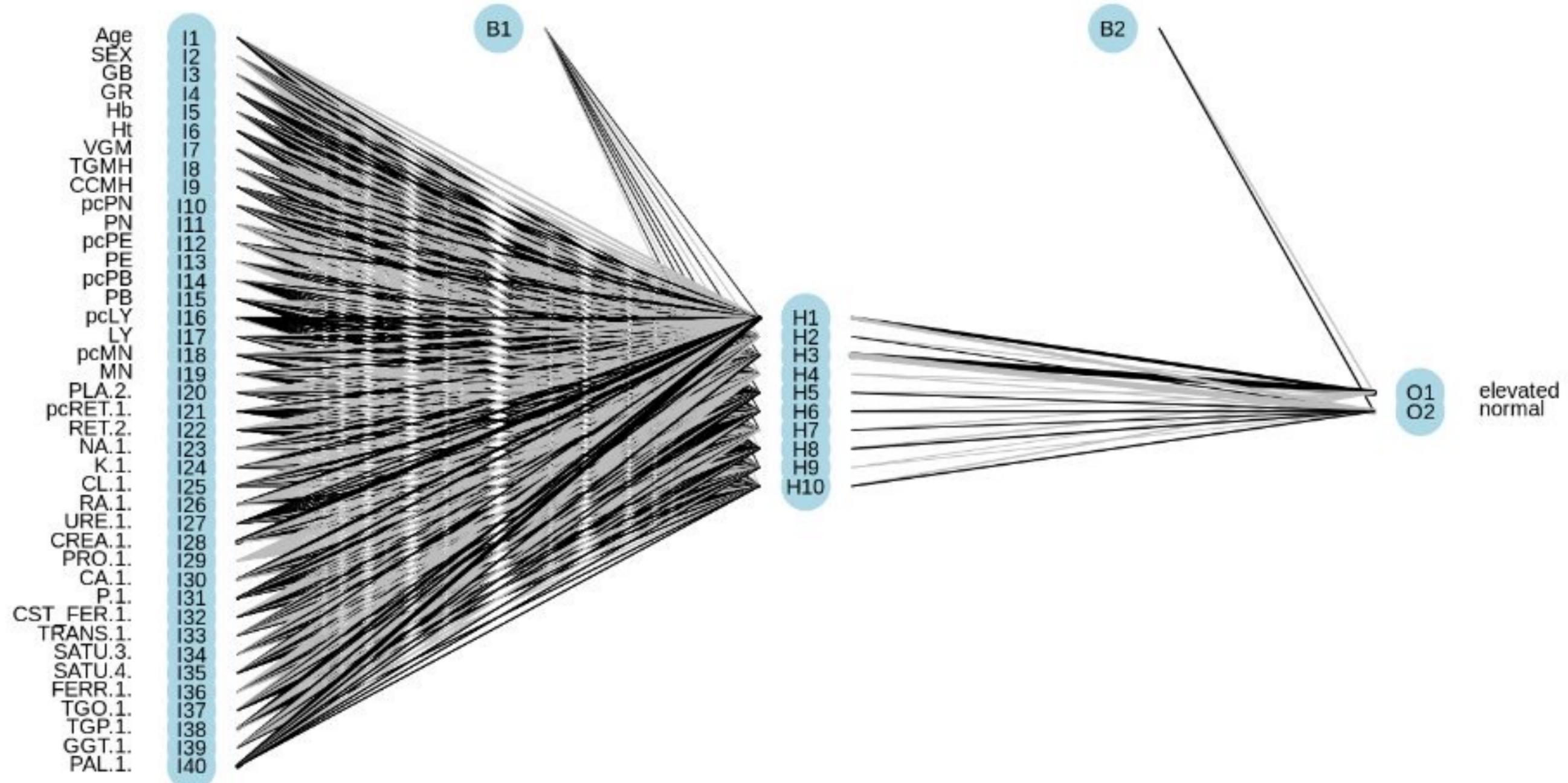
technique MSE



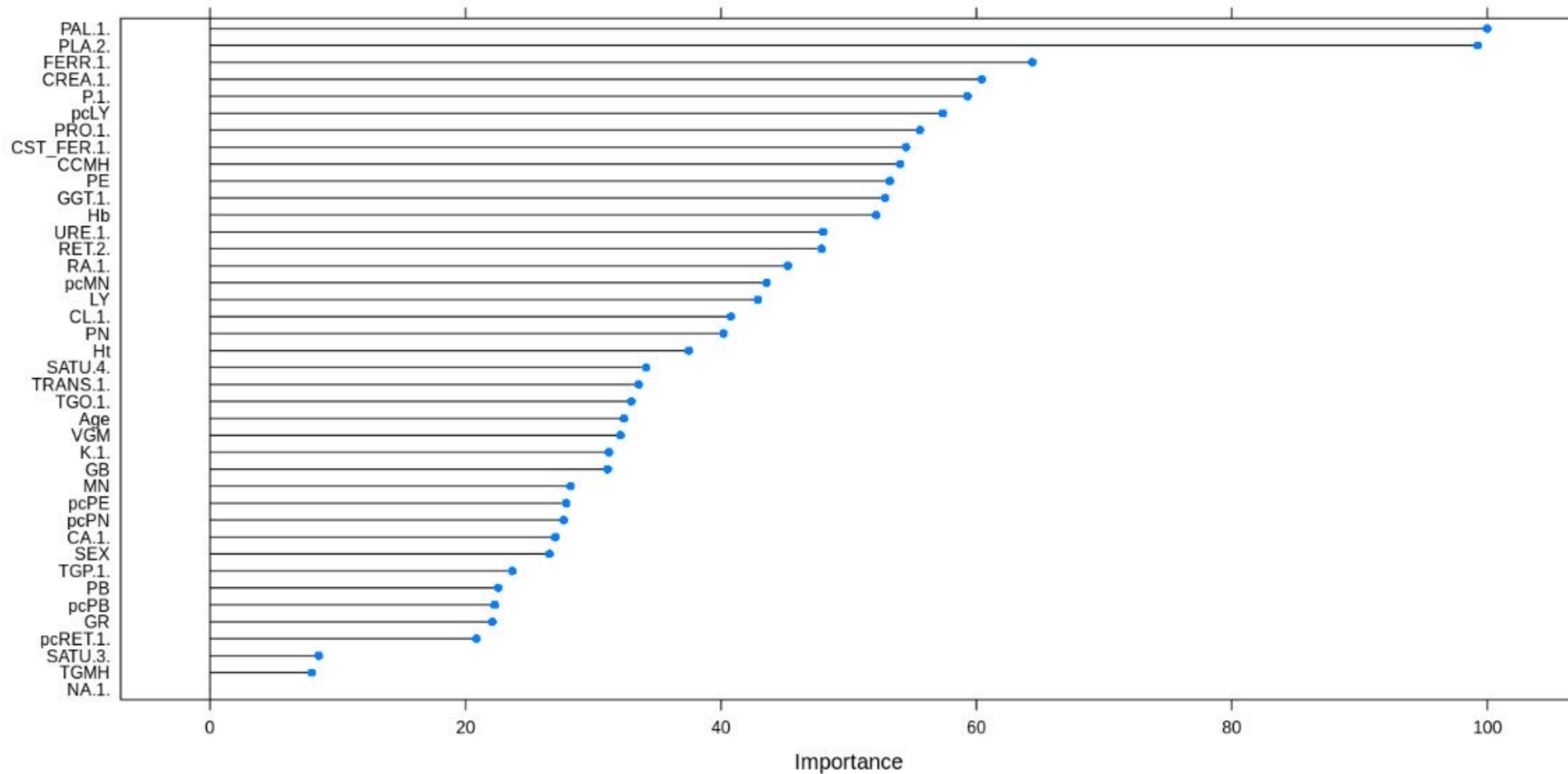
Fôrets Randomisées +++

Variables importantes selon le modèle
Forêts Randomisées

Résultats : Réseau de neurones



Résultats : Réseau de neurones



Discussion

Variables importantes

- Phosphatase alcaline, créatinine et NFS très important pour la prédiction Et réticulocytes +++
- Intérêts des modèles IA pour mettre en évidence des liens non connus+++ dans notre étude lien très marqué entre PTH et NFS++

Semblables aux autres études?

- *Machine Learning–Based Prediction of Elevated PTH Levels Among the US General Population*
- Hajime Kato, J Clin Endocrinol Metab 2022
- 8208 US adults (population générale)
753 (9.2%) showed PTH greater than 74 pg/mL
- **random forest AUC = 0.79** +++ (AUC comparable+++ 0.78)
GBM AUC = 0.78
Variables importantes : Créatinine, Age, Calcium, Urée,...

Cependant, ils n'ont pas utilisé NFS car non disponible dans données +++++

Réperer les variables intéressantes et facilement accessibles

Application du modèle en pratique

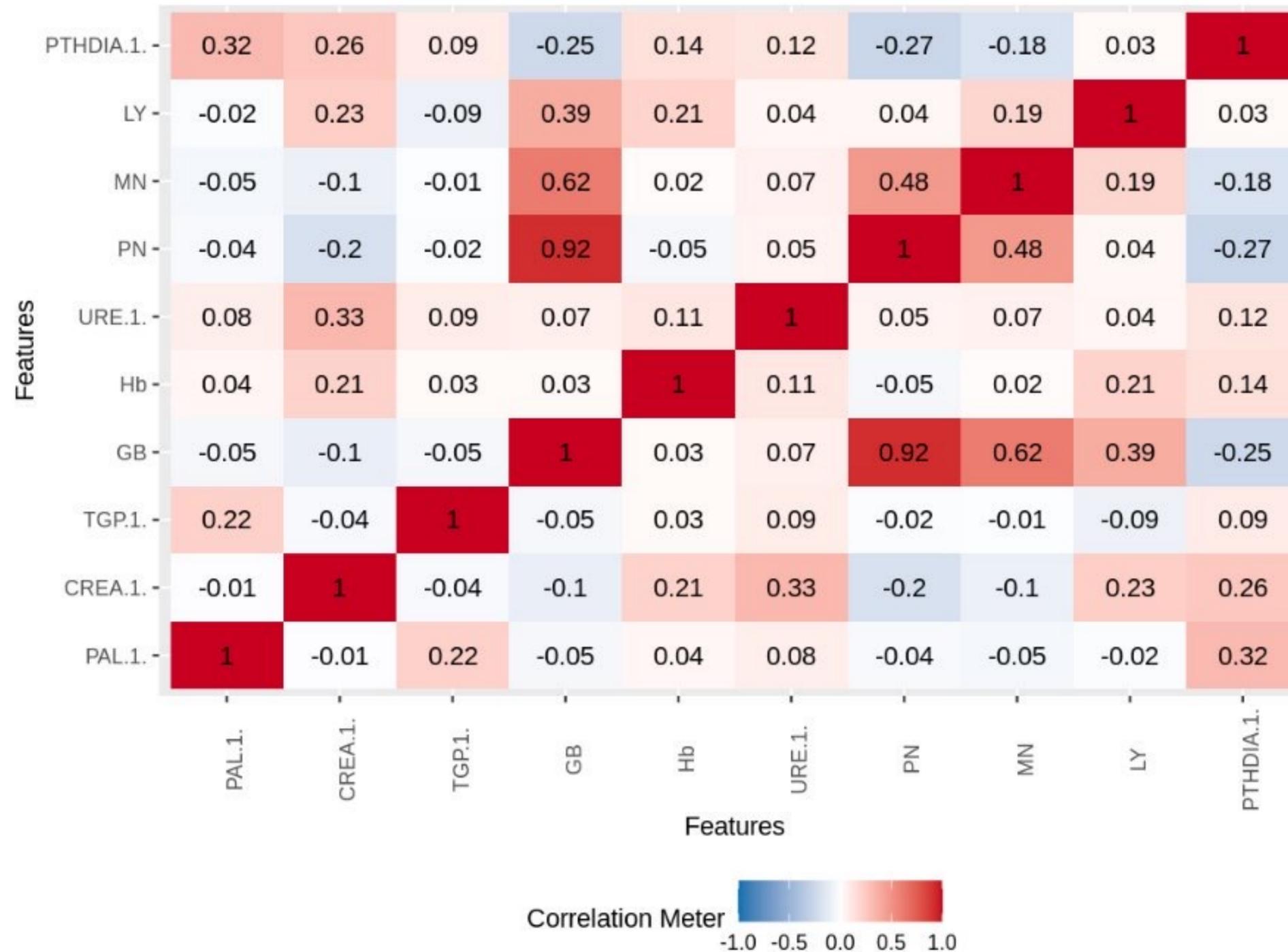
- Choisir des variables simples (paramètres faits en routine)
- Essayer au maximum d'exclure des paramètres difficiles (couteux+++ non fait en pratique+++)

-> **Permet d'appliquer les algorithmes sur des bilans simples de routines++**

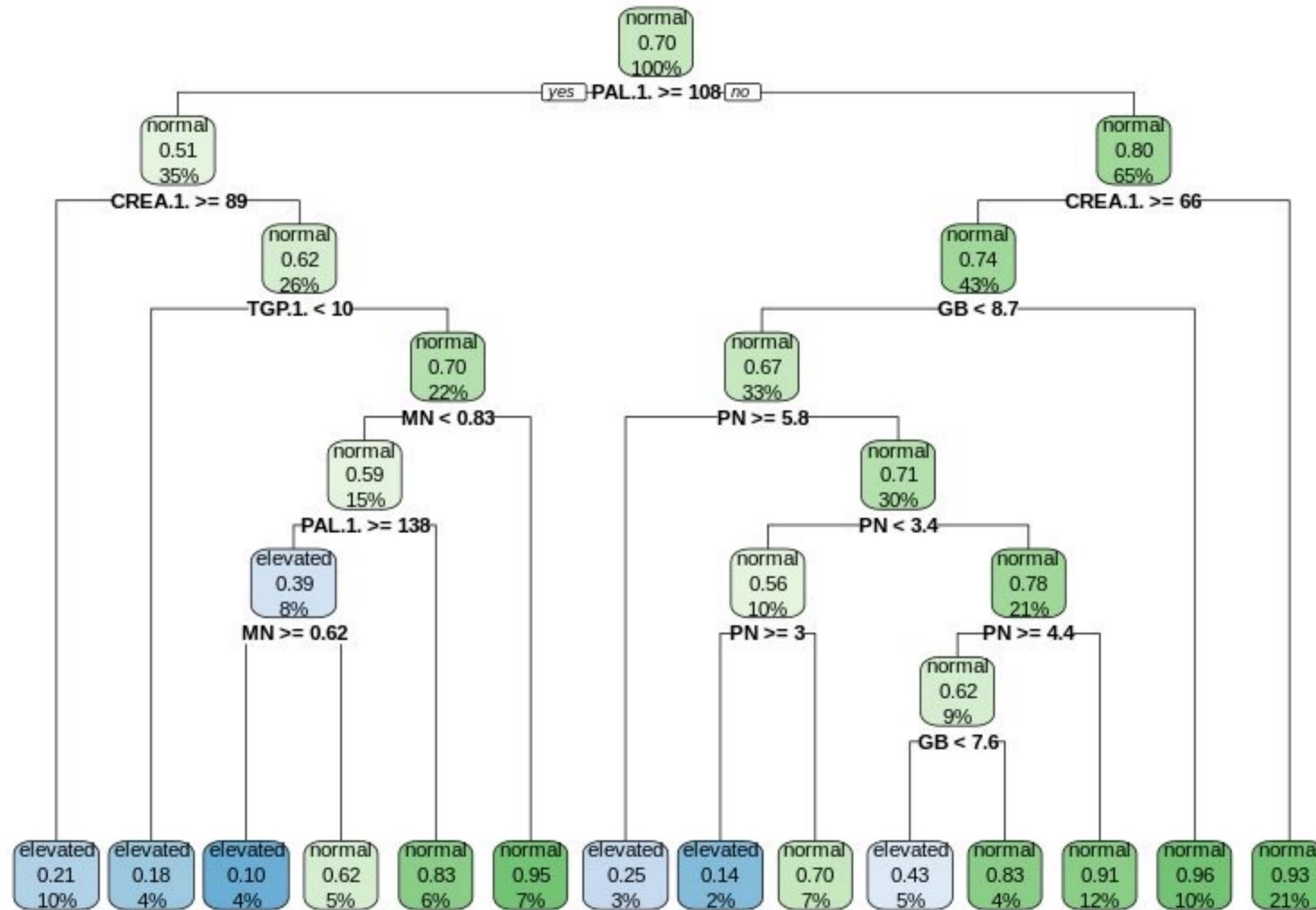
- Ici variable % réticulocyte, réticulocytes sont dosés 4 fois par an... ++++
- Nécessité de faire modèles en excluant réticulocytes

-> **Réentraînement du modèle en excluant réticulocytes**

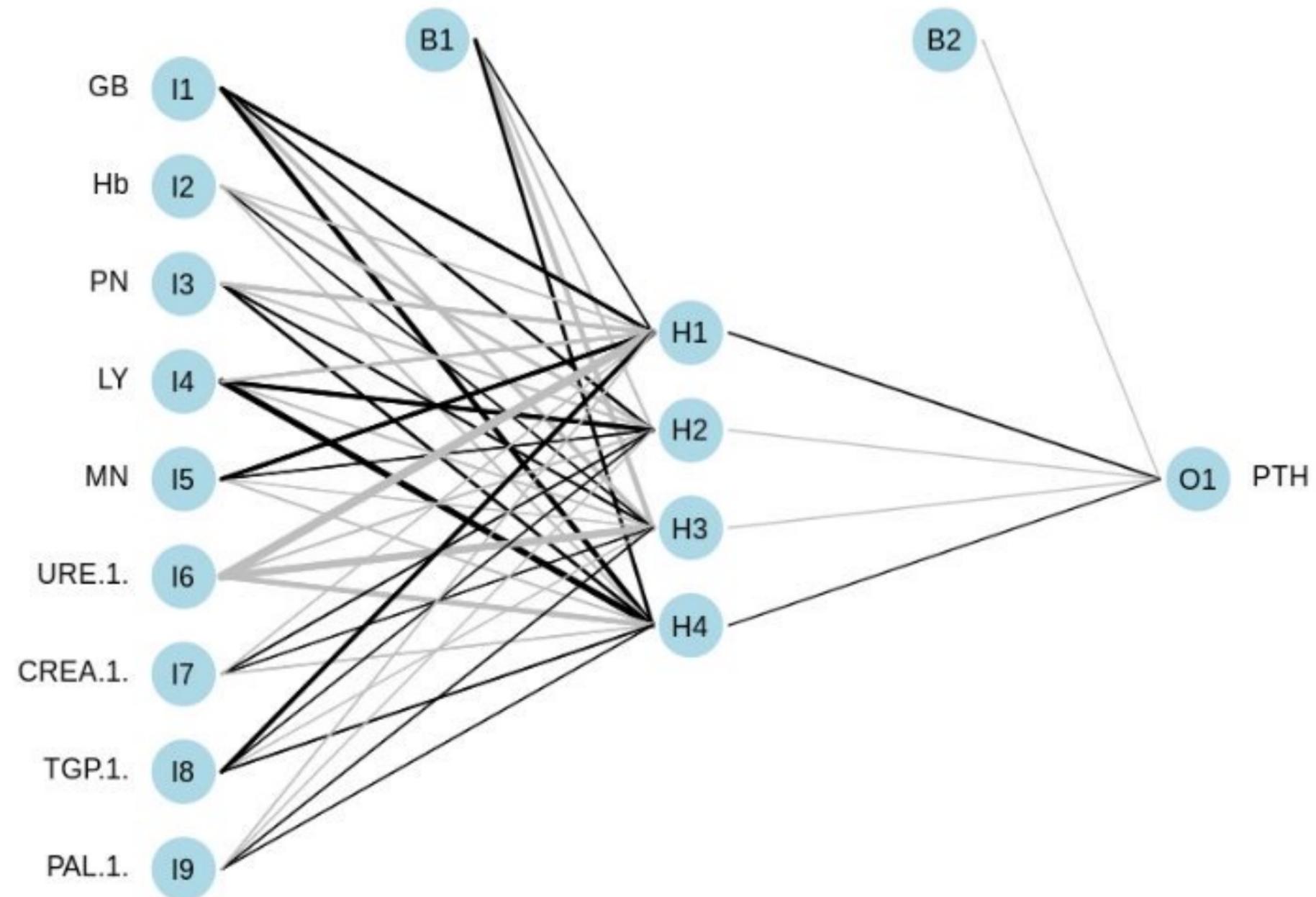
Discussion : Modèles simples "routine"



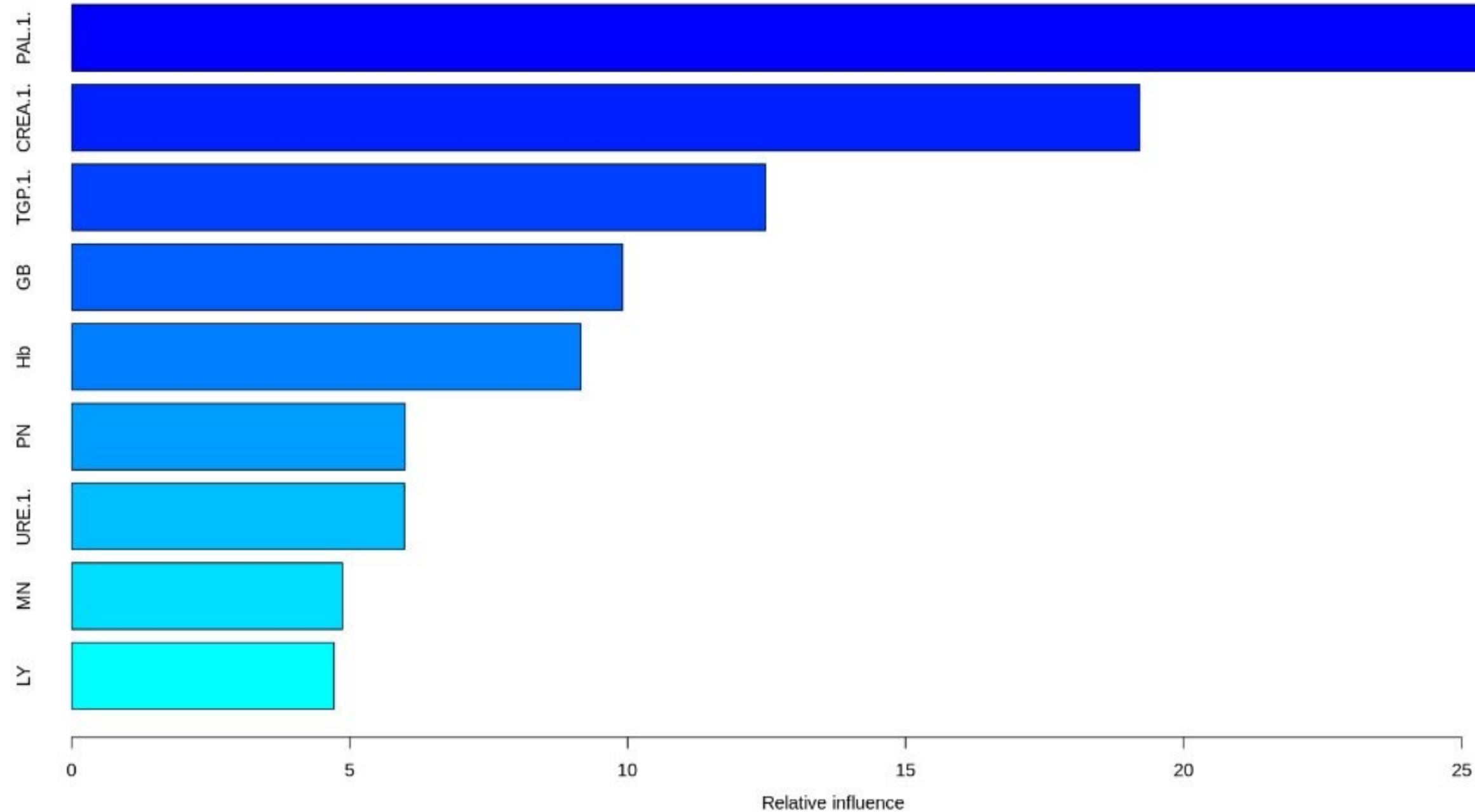
Discussion : Modèles simples "routine"



Discussion : Modèles simples "routine"

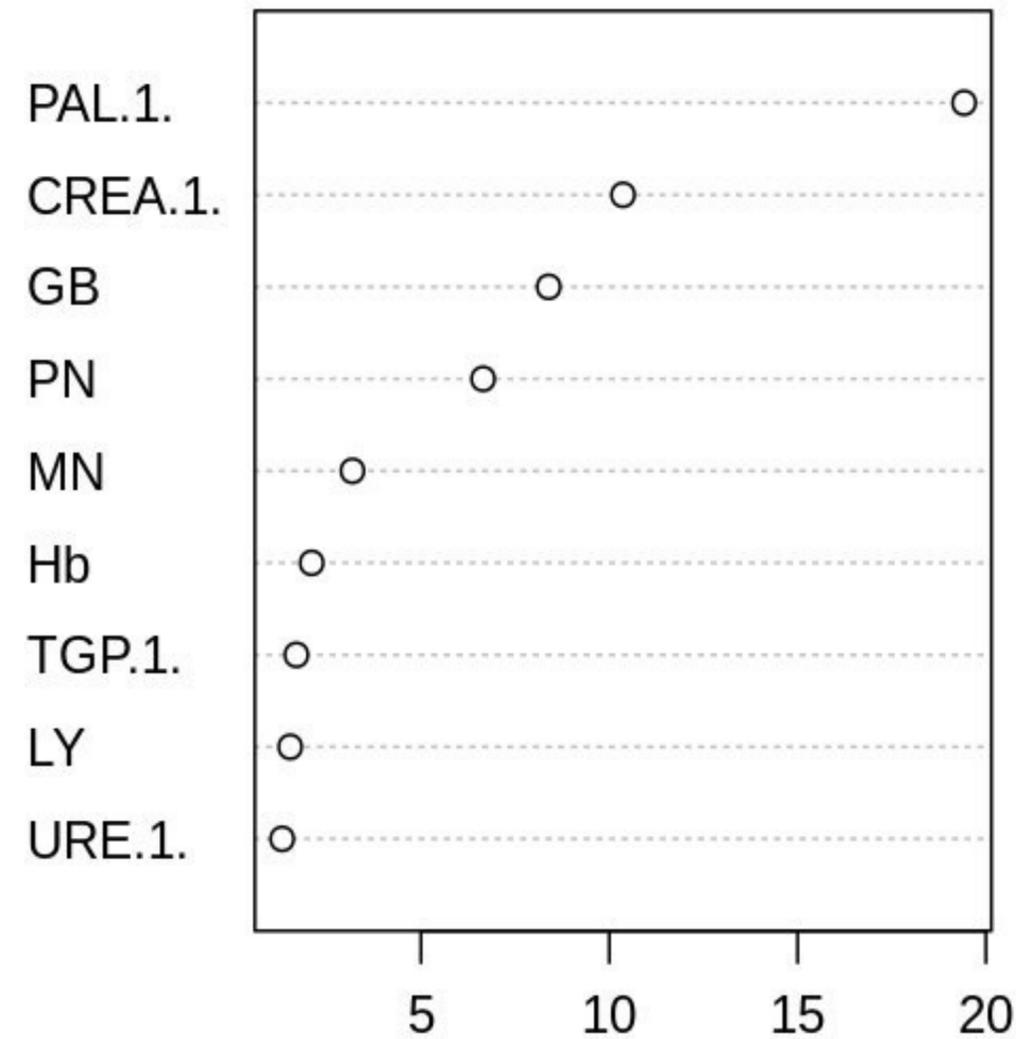


Discussion : Modèles simples "routine"



Importance des variables selon GBM

Discussion : Modèles simples "routine"



Importance des variables selon Fôrets Randomisées

Discussion : Modèles simples "routine"

Modèle	Arbre D	Arbre D routine	Fôrets	Forêts routine	NN	NN routine	GBM routine
Précision	75 %	65 %	88 %	75 %	62 %	65 %	62 %
Sensibilité	71 %	50 %	60 %	54 %	45 %	30 %	45 %
Spécificité	76 %	72 %	95 %	88 %	71 %	81%	71 %
VPP	45 %	45 %	75 %	75 %	45 %	42 %	45 %
VPN	90 %	76 %	90 %	75 %	71 %	72 %	71 %
AUC	73 %	61 %	78 %	71 %	58 %	56 %	58 %

Conclusion

Application

- Les patients présentant prédiction "PTH Elevée" pourront se voir proposer un dosage de la PTH
- Dépistage précoces des taux PTH hors cible KDIGO +++
- En pratique Rajout PTH si prediction "PTH élevée" , eg ECBU si BU positive.

Monitoring

- Surveillance Data long terme
- % Validation – Cellule qualité, IA Act, CORFRAC

Questions

Réponses

