

Design d'un algorithme d'IA en grande dimension pour prédire la réadmission à l'hôpital

Simon Bussy⁽¹⁾

simon.bussy@gmail.com

www.simonbussy.com



Étude réalisée avec

Raphaël Veil^(2,3), Vincent Looten^(2,3), Anita Burgun^(2,3), Stéphane Gaïffas^(1,4), Agathe Guilloux⁽⁵⁾, Brigitte Ranque^(6,7), Anne-Sophie Jannot^(2,3)

⁽¹⁾ LPSM, UMR 8001, Sorbonne University, Paris. ⁽²⁾ APHP, Département d'Informatique Biomédicale et de Santé Publique, HEGP, Paris. ⁽³⁾ INSERM, UMRS 1138, Eq22, Centre de Recherche des Cordeliers, Univ. Paris Descartes, Paris.

⁽⁴⁾ CMAP, UMR 7641, École Polytechnique CNRS, Paris. ⁽⁵⁾ LAMME, Univ. Evry, CNRS, Université Paris-Saclay, Paris.

⁽⁶⁾ INSERM, UMRS 970, Univ. Paris Descartes, Paris. ⁽⁷⁾ APHP, Département de Médecine Interne, HEGP, Paris.



Séminaire méthodologique sur les parcours de soins

Le 22/01/2019

- ▶ Prédiction du pronostic, médecine personnalisée

IA en santé

- ▶ Prédiction du pronostic, médecine personnalisée
- ▶ Données massives: dossiers médicaux électroniques (EHR)

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

IA en santé

- ▶ Prédiction du pronostic, médecine personnalisée
- ▶ Données massives: dossiers médicaux électroniques (EHR)
- ▶ Enjeux:

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

Introduction

IA en santé

- ▶ Prédiction du pronostic, médecine personnalisée
- ▶ Données massives: dossiers médicaux électroniques (EHR)
- ▶ Enjeux:
 - ▶ Modéliser ces données complexes

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

Introduction

IA en santé

- ▶ Prédiction du pronostic, médecine personnalisée
- ▶ Données massives: dossiers médicaux électroniques (EHR)
- ▶ Enjeux:
 - ▶ Modéliser ces données complexes
 - ▶ Élaborer un algorithme qui apprend et généralise

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

IA en santé

- ▶ Prédiction du pronostic, médecine personnalisée
- ▶ Données massives: dossiers médicaux électroniques (EHR)
- ▶ Enjeux:
 - ▶ Modéliser ces données complexes
 - ▶ Élaborer un algorithme qui apprend et généralise
 - ▶ Interprétabilité des résultats!

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

Introduction

IA en santé

- ▶ Prédiction du pronostic, médecine personnalisée
- ▶ Données massives: dossiers médicaux électroniques (EHR)
- ▶ Enjeux:
 - ▶ Modéliser ces données complexes
 - ▶ Élaborer un algorithme qui apprend et généralise
 - ▶ Interprétabilité des résultats!

Étude de cas

- ▶ Réadmission à l'hôpital après une crise vaso-occlusive (CVO) chez des patients atteints de drépanocytose

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

Introduction

IA en santé

- ▶ Prédiction du pronostic, médecine personnalisée
- ▶ Données massives: dossiers médicaux électroniques (EHR)
- ▶ Enjeux:
 - ▶ Modéliser ces données complexes
 - ▶ Élaborer un algorithme qui apprend et généralise
 - ▶ Interprétabilité des résultats!

Étude de cas

- ▶ Réadmission à l'hôpital après une crise vaso-occlusive (CVO) chez des patients atteints de drépanocytose
- ▶ But: Identifier

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

Introduction

IA en santé

- ▶ Prédiction du pronostic, médecine personnalisée
- ▶ Données massives: dossiers médicaux électroniques (EHR)
- ▶ Enjeux:
 - ▶ Modéliser ces données complexes
 - ▶ Élaborer un algorithme qui apprend et généralise
 - ▶ Interprétabilité des résultats!

Étude de cas

- ▶ Réadmission à l'hôpital après une crise vaso-occlusive (CVO) chez des patients atteints de drépanocytose
- ▶ But: Identifier
 - ▶ les patients à haut risque de réadmission précoce

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

Introduction

IA en santé

- ▶ Prédiction du pronostic, médecine personnalisée
- ▶ Données massives: dossiers médicaux électroniques (EHR)
- ▶ Enjeux:
 - ▶ Modéliser ces données complexes
 - ▶ Élaborer un algorithme qui apprend et généralise
 - ▶ Interprétabilité des résultats!

Étude de cas

- ▶ Réadmission à l'hôpital après une crise vaso-occlusive (CVO) chez des patients atteints de drépanocytose
- ▶ But: Identifier
 - ▶ les patients à haut risque de réadmission précoce
 - ▶ les facteurs prédictifs

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

I. Trajectories of biological values and vital parameters, a retrospective cohort study on non-complicated vaso-occlusive crises

Contexte

- ▶ Drépanocytose : maladie génétique la plus fréquente [10]

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

Contexte

- ▶ Drépanocytose : maladie génétique la plus fréquente [10]
- ▶ Crises vaso-occlusives (CVO)

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

Contexte

- ▶ Drépanocytose : maladie génétique la plus fréquente [10]
- ▶ Crises vaso-occlusives (CVO)
- ▶ Pas de biomarqueur pour le suivi

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

Contexte

- ▶ Drépanocytose : maladie génétique la plus fréquente [10]
- ▶ Crises vaso-occlusives (CVO)
- ▶ Pas de biomarqueur pour le suivi
- ▶ Étude rétrospective : cohorte de l'HEGP

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

Contexte

- ▶ Drépanocytose : maladie génétique la plus fréquente [10]
- ▶ Crises vaso-occlusives (CVO)
- ▶ Pas de biomarqueur pour le suivi
- ▶ Étude rétrospective : cohorte de l'HEGP

Objectifs

- ▶ Décrire l'évolution des biomarqueurs et paramètres vitaux lors d'une CVO "non compliquée"

Introduction

Contexte

- ▶ Drépanocytose : maladie génétique la plus fréquente [10]
- ▶ Crises vaso-occlusives (CVO)
- ▶ Pas de biomarqueur pour le suivi
- ▶ Étude rétrospective : cohorte de l'HEGP

Objectifs

- ▶ Décrire l'évolution des biomarqueurs et paramètres vitaux lors d'une CVO "non compliquée"
- ▶ Détecter la présence d'une complication

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

Contexte

- ▶ Drépanocytose : maladie génétique la plus fréquente [10]
- ▶ Crises vaso-occlusives (CVO)
- ▶ Pas de biomarqueur pour le suivi
- ▶ Étude rétrospective : cohorte de l'HEGP

Objectifs

- ▶ Décrire l'évolution des biomarqueurs et paramètres vitaux lors d'une CVO "non compliquée"
- ▶ Détecter la présence d'une complication
- ▶ Identifier quel(s) biomarqueur(s) surveiller

Données

- ▶ Patients admis à l'HEGP pour CVO entre 2010 → 2015

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

Données

- ▶ Patients admis à l'HEGP pour CVO entre 2010 → 2015
- ▶ 329 séjours, 164 patients

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

Méthode

Données

- ▶ Patients admis à l'HEGP pour CVO entre 2010 → 2015
- ▶ 329 séjours, 164 patients
- ▶ Données longitudinales

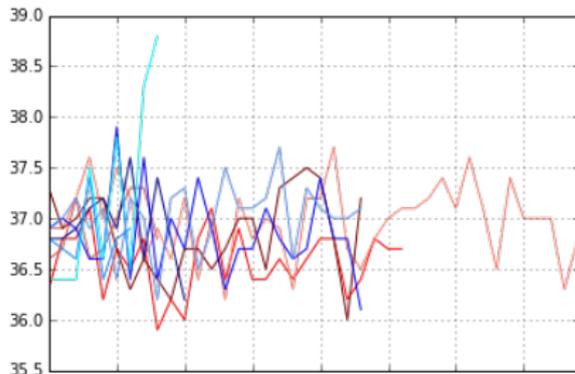


Figure 1: Évolution de la température pour 10 patients

Données

- ▶ Patients admis à l'HEGP pour CVO entre 2010 → 2015
- ▶ 329 séjours, 164 patients
- ▶ Données longitudinales

Description des trajectoires moyennes

Pour chaque trajectoire et pour chaque séjour i :

- ▶ on génère une grille uniforme de temps t_k

Données

- ▶ Patients admis à l'HEGP pour CVO entre 2010 → 2015
- ▶ 329 séjours, 164 patients
- ▶ Données longitudinales

Description des trajectoires moyennes

Pour chaque trajectoire et pour chaque séjour i :

- ▶ on génère une grille uniforme de temps t_k
- ▶ on ajuste un *spline* f_i

Données

- ▶ Patients admis à l'HEGP pour CVO entre 2010 → 2015
- ▶ 329 séjours, 164 patients
- ▶ Données longitudinales

Description des trajectoires moyennes

Pour chaque trajectoire et pour chaque séjour i :

- ▶ on génère une grille uniforme de temps t_k
- ▶ on ajuste un *spline* f_i
- ▶ on calcule $f_i(t_k)$ pour chaque temps t_k

Données

- ▶ Patients admis à l'HEGP pour CVO entre 2010 → 2015
- ▶ 329 séjours, 164 patients
- ▶ Données longitudinales

Description des trajectoires moyennes

Pour chaque trajectoire et pour chaque séjour i :

- ▶ on génère une grille uniforme de temps t_k
- ▶ on ajuste un *spline* f_i
- ▶ on calcule $f_i(t_k)$ pour chaque temps t_k
- ▶ on déduit une trajectoire moyenne avec IC

Exemples de résultats graphiques

Introduction

Trajectoires

Contexte

Méthode

Résultats

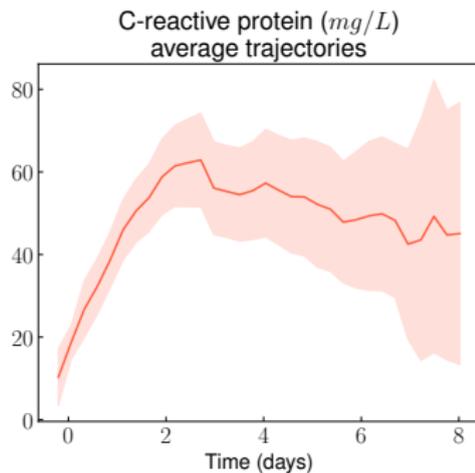
Réadmissions

Contexte

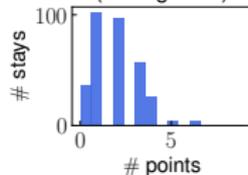
Méthode

Résultats

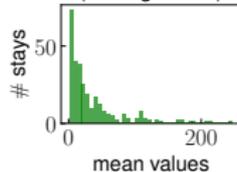
Références



Amount of measure points
(average=1.9)



Mean value distribution
(average=34.0)



Exemples de résultats graphiques

Introduction

Trajectoires

Contexte

Méthode

Résultats

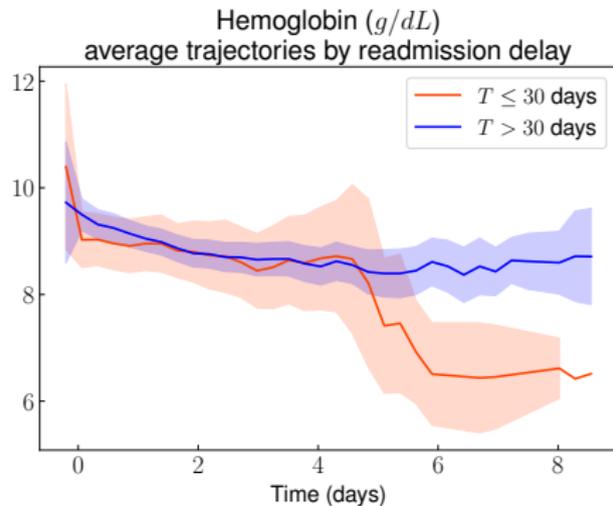
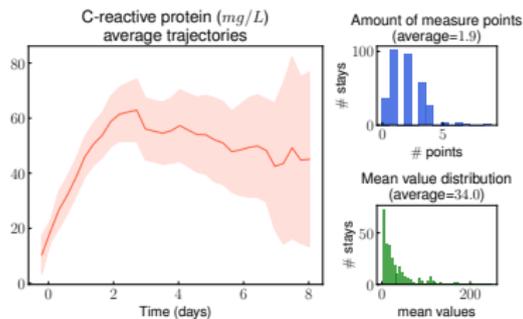
Réadmissions

Contexte

Méthode

Résultats

Références



- ▶ Plusieurs biomarqueurs et paramètres vitaux évoluent

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

- ▶ Plusieurs biomarqueurs et paramètres vitaux évoluent
- ▶ Extraction d'information pertinente à partir d'un entrepôt de données cliniques de grande dimension

Conclusion

- ▶ Plusieurs biomarqueurs et paramètres vitaux évoluent
- ▶ Extraction d'information pertinente à partir d'un entrepôt de données cliniques de grande dimension

Article associé

R. Veil, **S. Bussy**, V. Looten, J.B. Arlet, J. Pouchot Camoz, A.S. Jannot et B. Ranque

Trajectories of biological values and vital parameters : a retrospective cohort study on non-complicated vaso-occlusive crises

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

Conclusion

- ▶ Plusieurs biomarqueurs et paramètres vitaux évoluent
- ▶ Extraction d'information pertinente à partir d'un entrepôt de données cliniques de grande dimension

Article associé

R. Veil, **S. Bussy**, V. Looten, J.B. Arlet, J. Pouchot Camoz, A.S. Jannot et B. Ranque

Trajectories of biological values and vital parameters : a retrospective cohort study on non-complicated vaso-occlusive crises

Code Python

- ▶ Disponible à <https://github.com/SimonBussy/redcvo>

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

Conclusion

- ▶ Plusieurs biomarqueurs et paramètres vitaux évoluent
- ▶ Extraction d'information pertinente à partir d'un entrepôt de données cliniques de grande dimension

Article associé

R. Veil, **S. Bussy**, V. Looten, J.B. Arlet, J. Pouchot Camoz, A.S. Jannot et B. Ranque

Trajectories of biological values and vital parameters : a retrospective cohort study on non-complicated vaso-occlusive crises

Code Python

- ▶ Disponible à <https://github.com/SimonBussy/redcvo>
- ▶ Figures pour tous les biomarqueurs et paramètres vitaux

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

[Introduction](#)[Trajectoires](#)[Contexte](#)[Méthode](#)[Résultats](#)[Réadmissions](#)[Contexte](#)[Méthode](#)[Résultats](#)[Références](#)

II. Early-readmission prediction in a high-dimensional heterogeneous covariates and time-to-event outcome framework

- ▶ Facteurs prédictifs de la réhospitalisation pour CVO

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

- ▶ Facteurs prédictifs de la réhospitalisation pour CVO
- ▶ Problèmes récurrents dans les études cliniques :

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

- ▶ Facteurs prédictifs de la réhospitalisation pour CVO
- ▶ Problèmes récurrents dans les études cliniques :
 - ▶ prédiction du risque

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

- ▶ Facteurs prédictifs de la réhospitalisation pour CVO
- ▶ Problèmes récurrents dans les études cliniques :
 - ▶ prédiction du risque
 - ▶ identification des covariables impliquées

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

Introduction

- ▶ Facteurs prédictifs de la réhospitalisation pour CVO
- ▶ Problèmes récurrents dans les études cliniques :
 - ▶ prédiction du risque
 - ▶ identification des covariables impliquées

Deux cadres

- ▶ *Binary outcome setting*

Introduction

- ▶ Facteurs prédictifs de la réhospitalisation pour CVO
- ▶ Problèmes récurrents dans les études cliniques :
 - ▶ prédiction du risque
 - ▶ identification des covariables impliquées

Deux cadres

- ▶ *Binary outcome setting*
 - ▶ Réhospitalisation “précoce” en se basant sur un seuil ϵ pré-défini

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

Introduction

- ▶ Facteurs prédictifs de la réhospitalisation pour CVO
- ▶ Problèmes récurrents dans les études cliniques :
 - ▶ prédiction du risque
 - ▶ identification des covariables impliquées

Deux cadres

- ▶ *Binary outcome setting*
 - ▶ Réhospitalisation “précoce” en se basant sur un seuil ϵ pré-défini
 - ▶ Résultats très dépendants du choix de ϵ [4]

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

Introduction

- ▶ Facteurs prédictifs de la réhospitalisation pour CVO
- ▶ Problèmes récurrents dans les études cliniques :
 - ▶ prédiction du risque
 - ▶ identification des covariables impliquées

Deux cadres

- ▶ *Binary outcome setting*
 - ▶ Réhospitalisation “précoce” en se basant sur un seuil ϵ pré-défini
 - ▶ Résultats très dépendants du choix de ϵ [4]
- ▶ *Survival analysis setting* : pas de seuil *a priori*

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

Introduction

- ▶ Facteurs prédictifs de la réhospitalisation pour CVO
- ▶ Problèmes récurrents dans les études cliniques :
 - ▶ prédiction du risque
 - ▶ identification des covariables impliquées

Deux cadres

- ▶ *Binary outcome setting*
 - ▶ Réhospitalisation “précoce” en se basant sur un seuil ϵ pré-défini
 - ▶ Résultats très dépendants du choix de ϵ [4]
- ▶ *Survival analysis setting* : pas de seuil *a priori*

Objectif

Comparer des méthodes d'apprentissage issues de ces deux cadres

- ▶ Étude rétrospective monocentrique sur la cohorte de I.

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

- ▶ Étude rétrospective monocentrique sur la cohorte de I.
- ▶ On tire aléatoirement 1 séjour par patient (*i.i.d.*)

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

Cas d'étude

- ▶ Étude rétrospective monocentrique sur la cohorte de I.
- ▶ On tire aléatoirement 1 séjour par patient (*i.i.d.*)

Extraction de covariables

À partir des données longitudinales, par exemple :

- ▶ Pente d'une régression linéaire (dernières 48h)

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

Cas d'étude

- ▶ Étude rétrospective monocentrique sur la cohorte de I.
- ▶ On tire aléatoirement 1 séjour par patient (*i.i.d.*)

Extraction de covariables

À partir des données longitudinales, par exemple :

- ▶ Pente d'une régression linéaire (dernières 48h)
- ▶ Hyper-paramètres des noyaux de PG [11]

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

Cas d'étude

- ▶ Étude rétrospective monocentrique sur la cohorte de I.
- ▶ On tire aléatoirement 1 séjour par patient (*i.i.d.*)

Extraction de covariables

À partir des données longitudinales, par exemple :

- ▶ Pente d'une régression linéaire (dernières 48h)
- ▶ Hyper-paramètres des noyaux de PG [11]

174 covariables

- ▶ démographiques

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

Cas d'étude

- ▶ Étude rétrospective monocentrique sur la cohorte de I.
- ▶ On tire aléatoirement 1 séjour par patient (*i.i.d.*)

Extraction de covariables

À partir des données longitudinales, par exemple :

- ▶ Pente d'une régression linéaire (dernières 48h)
- ▶ Hyper-paramètres des noyaux de PG [11]

174 covariables

- ▶ démographiques
- ▶ qualitatives (ex: type d'opioïde administré)

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

Cas d'étude

- ▶ Étude rétrospective monocentrique sur la cohorte de I.
- ▶ On tire aléatoirement 1 séjour par patient (*i.i.d.*)

Extraction de covariables

À partir des données longitudinales, par exemple :

- ▶ Pente d'une régression linéaire (dernières 48h)
- ▶ Hyper-paramètres des noyaux de PG [11]

174 covariables

- ▶ démographiques
- ▶ qualitatives (ex: type d'opioïde administré)
- ▶ quantitative (ex: paramètres biologiques/vitaux/doses d'opioïde)

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

Modèles considérés

- ▶ *Binary outcome setting* ($\epsilon = 30$ jours), score : AUC [1]

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

Modèles considérés

- ▶ *Binary outcome setting* ($\epsilon = 30$ jours), score : AUC [1]
 - ▶ Régression Logistique (LR) [9]

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

Modèles considérés

- ▶ *Binary outcome setting* ($\epsilon = 30$ jours), score : AUC [1]
 - ▶ Régression Logistique (LR) [9]
 - ▶ SVM avec noyau linéaire [12]

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

Modèles considérés

- ▶ *Binary outcome setting* ($\epsilon = 30$ jours), score : AUC [1]
 - ▶ Régression Logistique (LR) [9]
 - ▶ SVM avec noyau linéaire [12]
 - ▶ Forêts aléatoires (RF) [2]

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

Modèles considérés

- ▶ *Binary outcome setting* ($\epsilon = 30$ jours), score : AUC [1]
 - ▶ Régression Logistique (LR) [9]
 - ▶ SVM avec noyau linéaire [12]
 - ▶ Forêts aléatoires (RF) [2]
 - ▶ Gradient Boosting (GB) [7]

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

Modèles considérés

- ▶ *Binary outcome setting* ($\epsilon = 30$ jours), score : AUC [1]
 - ▶ Régression Logistique (LR) [9]
 - ▶ SVM avec noyau linéaire [12]
 - ▶ Forêts aléatoires (RF) [2]
 - ▶ Gradient Boosting (GB) [7]
 - ▶ Réseaux de neurones à une couche cachée (NN) [13]

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

Modèles considérés

- ▶ *Binary outcome setting* ($\epsilon = 30$ jours), score : AUC [1]
 - ▶ Régression Logistique (LR) [9]
 - ▶ SVM avec noyau linéaire [12]
 - ▶ Forêts aléatoires (RF) [2]
 - ▶ Gradient Boosting (GB) [7]
 - ▶ Réseaux de neurones à une couche cachée (NN) [13]
- ▶ *Survival analysis setting*, score : C-index [8]

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

Modèles considérés

- ▶ *Binary outcome setting* ($\epsilon = 30$ jours), score : AUC [1]
 - ▶ Régression Logistique (LR) [9]
 - ▶ SVM avec noyau linéaire [12]
 - ▶ Forêts aléatoires (RF) [2]
 - ▶ Gradient Boosting (GB) [7]
 - ▶ Réseaux de neurones à une couche cachée (NN) [13]
- ▶ *Survival analysis setting*, score : C-index [8]
 - ▶ Cox PH [5]

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

Modèles considérés

- ▶ *Binary outcome setting* ($\epsilon = 30$ jours), score : AUC [1]
 - ▶ Régression Logistique (LR) [9]
 - ▶ SVM avec noyau linéaire [12]
 - ▶ Forêts aléatoires (RF) [2]
 - ▶ Gradient Boosting (GB) [7]
 - ▶ Réseaux de neurones à une couche cachée (NN) [13]
- ▶ *Survival analysis setting*, score : C-index [8]
 - ▶ Cox PH [5]
 - ▶ CURE [6]

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

Modèles considérés

- ▶ *Binary outcome setting* ($\epsilon = 30$ jours), score : AUC [1]
 - ▶ Régression Logistique (LR) [9]
 - ▶ SVM avec noyau linéaire [12]
 - ▶ Forêts aléatoires (RF) [2]
 - ▶ Gradient Boosting (GB) [7]
 - ▶ Réseaux de neurones à une couche cachée (NN) [13]
- ▶ *Survival analysis setting*, score : C-index [8]
 - ▶ Cox PH [5]
 - ▶ CURE [6]
 - ▶ C-mix [3]

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références

Méthode

Modèles considérés

- ▶ *Binary outcome setting* ($\epsilon = 30$ jours), score : AUC [1]
 - ▶ Régression Logistique (LR) [9]
 - ▶ SVM avec noyau linéaire [12]
 - ▶ Forêts aléatoires (RF) [2]
 - ▶ Gradient Boosting (GB) [7]
 - ▶ Réseaux de neurones à une couche cachée (NN) [13]
- ▶ *Survival analysis setting*, score : C-index [8]
 - ▶ Cox PH [5]
 - ▶ CURE [6]
 - ▶ C-mix [3]

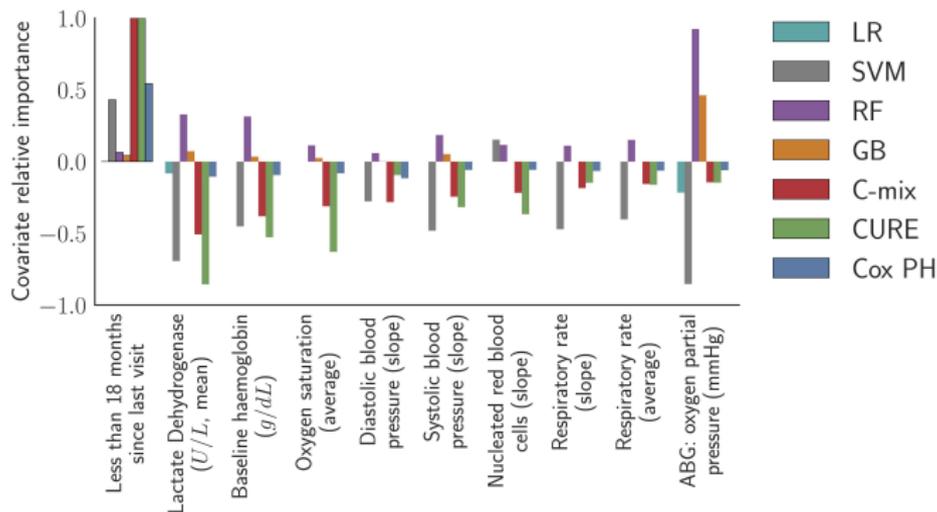
Comparaison des deux cadres

Prédiction de $T_i > \epsilon$ par $\hat{S}_i(\epsilon | X_i = x_i)$, score : AUC

Table 1: Performances en prédiction

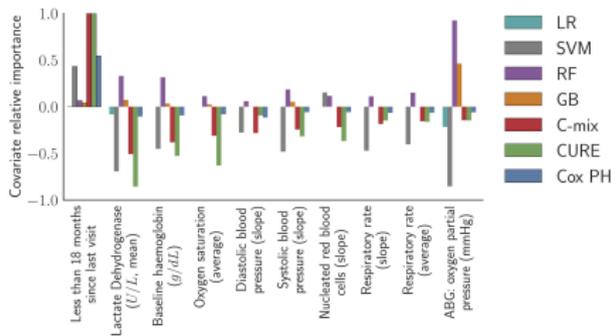
Setting	Métrique	Modèle	Score
Survival analysis	C-index	CURE	0.718
		Cox PH	0.725
		C-mix	0.754
Binary outcome ($\epsilon = 30$)	AUC	SVM	0.524
		GB	0.561
		LR	0.616
		NN	0.707
		RF	0.738
		$\hat{S}^{\text{CURE}}(\epsilon)$	0.831
		$\hat{S}^{\text{Cox PH}}(\epsilon)$	0.855
$\hat{S}^{\text{C-mix}}(\epsilon)$	0.940		

Résultats en sélection de variables

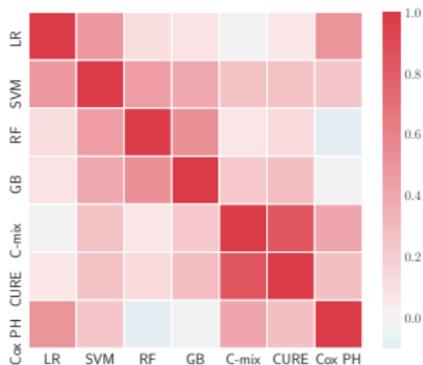


(a) Importance des 10 covariables les plus prédictives

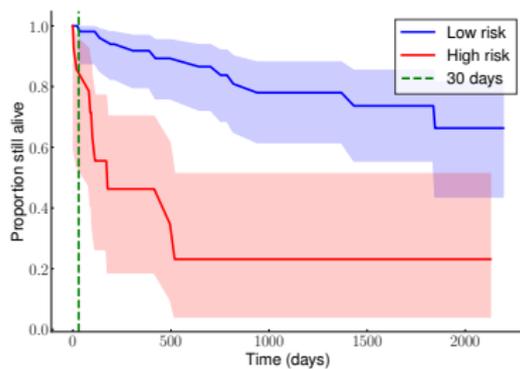
Résultats en sélection de variables



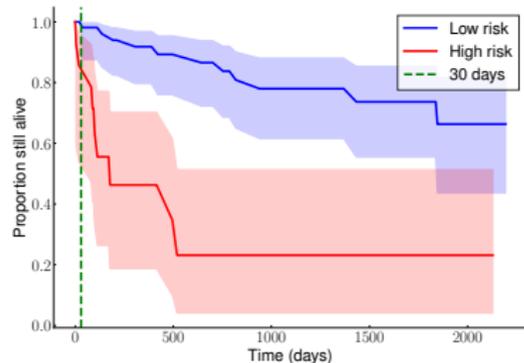
(a) Importance des 10 covariables les plus prédictives



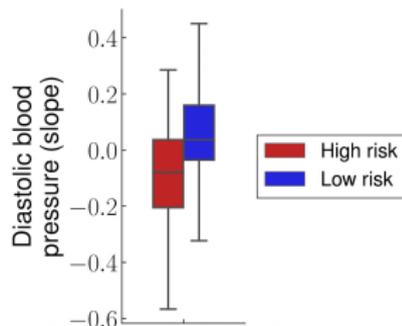
(b) Corrélation linéaire en sélection de variables



(a) Courbes de survie



(a) Courbes de survie



(b) Boxplot

- ▶ Modèles d'analyse de survie → fonctions de survie estimées → prédictions binaires pour un ϵ donné

Conclusion

- ▶ Modèles d'analyse de survie → fonctions de survie estimées → prédictions binaires pour un ϵ donné
- ▶ Méthodologie pour la création de features pertinentes

- ▶ Modèles d'analyse de survie → fonctions de survie estimées → prédictions binaires pour un ϵ donné
- ▶ Méthodologie pour la création de features pertinentes
- ▶ Bonnes performances du C-mix + aspects d'interprétation intéressants

Conclusion

- ▶ Modèles d'analyse de survie → fonctions de survie estimées → prédictions binaires pour un ϵ donné
- ▶ Méthodologie pour la création de features pertinentes
- ▶ Bonnes performances du C-mix + aspects d'interprétation intéressants

Article associé

S. Bussy, R. Veil, V. Looten, A. Burgun, S. Gaïffas, A. Guilloux, B. Ranque et A.S. Jannot

Comparison of methods for early-readmission prediction in a high dimensional heterogeneous covariates and time-to-event outcome framework

Accepté à *BMC Medical Research Methodology*, 2018.

Conclusion

- ▶ Modèles d'analyse de survie \rightarrow fonctions de survie estimées \rightarrow prédictions binaires pour un ϵ donné
- ▶ Méthodologie pour la création de features pertinentes
- ▶ Bonnes performances du C-mix + aspects d'interprétation intéressants

Article associé

S. Bussy, R. Veil, V. Looten, A. Burgun, S. Gaïffas, A. Guilloux, B. Ranque et A.S. Jannot

Comparison of methods for early-readmission prediction in a high dimensional heterogeneous covariates and time-to-event outcome framework

Accepté à *BMC Medical Research Methodology*, 2018.

Code Python

Disponible à

<https://github.com/SimonBussy/early-readmission-prediction>

- [1] Andrew P Bradley. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern recognition*, 30(7):1145–1159, 1997.
- [2] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- [3] Simon Bussy, Agathe Guilloux, Stéphane Gaïffas, and Anne-Sophie Jannot. C-mix: A high-dimensional mixture model for censored durations, with applications to genetic data. *Statistical Methods in Medical Research*, 0(0):0962280218766389, 2018.
- [4] Hung-Chia Chen, Ralph L Kodell, Kuang Fu Cheng, and James J Chen. Assessment of performance of survival prediction models for cancer prognosis. *BMC medical research methodology*, 12(1):102, 2012.
- [5] David R Cox. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220, 1972.
- [6] Vern T Farewell. The use of mixture models for the analysis of sureval data with long-term survivors. *Biometrics*, 38(4):1041–1046, 1982.
- [7] Jerome H Friedman. Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38(4):367–378, 2002.
- [8] F. E. Harrell, K. L. Lee, and D. B. Mark. Tutorial in biostatistics multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in medicine*, 15:361–387, 1996.
- [9] David W Hosmer Jr, Stanley Lemeshow, and Rodney X Sturdivant. *Applied logistic regression*, volume 398. John Wiley & Sons, 2013.
- [10] Frédéric B Piel, Anand P Patil, Rosalind E Howes, Oscar A Nyangiri, Peter W Gething, Mewahyu Dewi, William H Temperley, Thomas N Williams, David J Weatherall, and Simon I Hay. Global epidemiology of sickle haemoglobin in neonates: a contemporary geostatistical model-based map and population estimates. *The Lancet*, 381(9861):142–151, 2013.
- [11] Marco Pimentel, David A Clifton, Lei Clifton, and Lionel Tarassenko. Modelling patient time-series data from electronic health records using gaussian processes. In *Advances in Neural Information Processing Systems: Workshop on Machine Learning for Clinical Data Analysis*, pages 1–4, 2013.
- [12] Bernhard Schölkopf and Alexander J Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002.
- [13] B Yegnanarayana. *Artificial neural networks*. PHI Learning Pvt. Ltd., 2009.

Simon Bussy

Introduction

Trajectoires

Contexte

Méthode

Résultats

Réadmissions

Contexte

Méthode

Résultats

Références