

# Réseaux Bayésiens pour la Prédiction d'Évènements de Santé Défavorables en EHPAD (Établissement d'Hébergement pour Personnes Âgées Dépendantes)

Clara Charon<sup>1,3</sup>, Pierre-Henri Wuillemin<sup>1</sup>, Joël Belmin<sup>2,4</sup>

<sup>1</sup> LIP6 (UMR 7606), Sorbonne Université, 4 place Jussieu, 75005 Paris, France

<sup>2</sup> AP-HP, Hôpital Charles-Foix, Ivry-sur-Seine, France

<sup>3</sup> Teranga Software, 21 rue Viète, Paris, France

<sup>4</sup> Sorbonne Université, Paris, France

clara.charon@lip6.fr, pierre-henri.wuillemin@lip6.fr, j.belmin@aphp.fr

## Abstract

Cette étude propose l'utilisation de réseaux bayésiens avec la librairie Python `pyAgrum`<sup>1</sup> pour la prédiction d'évènements de santé défavorables, et plus particulièrement les escarres, en EHPAD. Après un prétraitement complexe, dû à une base de données événementielle, nous obtenons un classifieur intéressant et pertinent, que nous avons pu confronter à un avis d'expert et qui est meilleur que les méthodes actuellement utilisés en EHPAD.

**Mots-clés :** *Machine Learning, Réseaux bayésiens, Escarres, EHPAD*

## Introduction

L'EHPAD, établissement d'hébergement pour personnes âgées dépendantes, est une solution de plus en plus privilégiée face au vieillissement de la population, réel enjeu de santé publique actuel. La moyenne d'âge des résidents aujourd'hui est à 87 ans et la durée moyenne de séjour de 2,5 ans (Belmin et al. 2016). Ces établissements hébergent des personnes qui ne peuvent plus être prises en charge à leur domicile et leur assure une aide pour les gestes de la vie quotidienne.

Avec le développement des nouvelles technologies informatiques dans le domaine de la santé, un nombre croissant d'établissements de santé sont équipés de systèmes d'information regroupant les données administratives et médicales des patients ainsi que des informations sur les soins qui leur sont prodigués. Ces systèmes, et notamment les dossiers patients informatisés, permettent un accès facile et rapide aux informations des patients notamment dans un but d'amélioration de la qualité et de la sécurité des soins.

Le but de ce projet est d'utiliser les données de ses logiciels de dossier patient en EHPAD pour chercher des algorithmes capables de prédire plusieurs évènements de santé

Copyright © 2021, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

<sup>1</sup>`pyAgrum` est un wrapper Python pour le framework `aGrUM` (A GRaphical Universal Modeler) qui est une librairie C++ open source permettant la manipulation (apprentissage, modélisation, inférence) de données probabilistes : <http://agrum.gitlab.io>

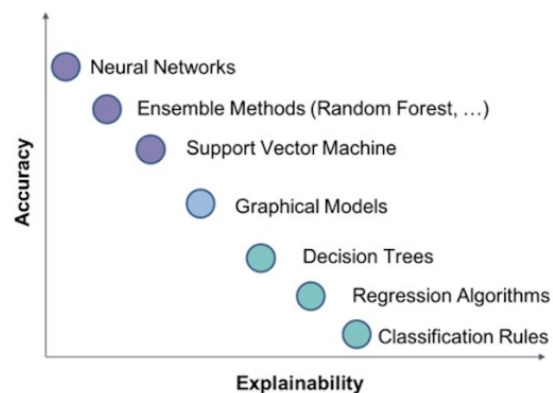


Figure 1: Classification des algorithmes de *Machine Learning* en fonction de leur pouvoir prédictif et de leur explicabilité (Dam, Tran, and Ghose 2018).

défavorables qui sont potentiellement modifiables par des interventions de santé appropriées. On se concentre particulièrement sur les méthodes capables de générer des explications comme les modèles graphiques probabilistes (Koller and Friedman 2009) et plus précisément les réseaux bayésiens (Pearl 1988).

Comme le montre la figure 1, utiliser les modèles graphiques est un compromis. En effet, nous faisons le choix de peut-être perdre en pouvoir prédictif, mais de gagner en explication, pour pouvoir apporter de la connaissance d'un point de vue médical (comme de nouveaux facteurs de risque). Cela permet aussi de donner confiance au personnel soignant qui va l'utiliser car il comprend le modèle et la décision de l'algorithme.

Le principe est donc de récupérer un échantillon conséquent de patients avec un maximum d'informations sur eux (prescriptions, pathologies, évaluations, habitudes de vie/alimentaires, etc...) et dont on sait s'ils ont subi l'évènement ciblé et quand. On constitue alors une base d'apprentissage avec les variables des patients une certaine période avant la survenue de l'évènement défavorable et

## Échelle de Braden (www.escarre.fr)

Sensibilité	Humidité	Activité
1 complétement limitée	1 constamment humide	1 confiné au lit
2 très limitée	2 très humide	2 confiné en chaise
3 légèrement limitée	3 parfois humide	3 marche parfois
4 pas de gêne	4 rarement humide	4 marche fréquemment
Mobilité	Nutrition	Friction et frottements
1 totalement immobile	1 très pauvre	1 problème permanent
2 très limitée	2 probablement inadéquate	2 problème potentiel
3 légèrement limitée	3 correcte	3 pas de problème
4 pas de limitation	4 excellente	

≥ 18 : risque bas 13 à 17 : risque modéré 8 à 12 : risque élevé ≤ 7 : risque élevé

Figure 2: Calcul de l'échelle de Braden.

des patients témoins. Deux classes sont définies : va subir l'évènement ou non. L'un des enjeux importants est la prise en compte de la transformation d'une base de donnée événementielle en données utilisables par un apprentissage statistique, tout en gardant un sens médical.

Dans un premier temps, la prédiction est limitée à la survenue de la première escarre d'un résident dans un établissement. Nous allons donc définir l'escarre, les méthodes actuellement utilisées pour les détecter et la façon dont nous allons évaluer notre modèle. Puis, nous aborderons le prétraitement et l'apprentissage de notre classifieur automatique. Enfin, nous montrerons les résultats numériques et graphiques obtenus.

## Prise en compte d'un évènement de santé défavorable : l'escarre

### L'escarre

Une escarre est une lésion cutanée d'origine ischémique (diminution de l'apport sanguin artériel à un organe) localisée au niveau de la peau ou des tissus mous sous-jacents, située en général sur une saillie osseuse. Elle est la conséquence d'un phénomène de pression ou de pression associée à un cisaillement.

C'est une pathologie humiliante pour le patient, douloureuse, qui diminue significativement la qualité de vie. Elle est aussi difficile et longue à traiter et le coût de cicatrization est non-négligeable et augmente exponentiellement avec le stade de la maladie (HAS 2001).

Les escarres sont pourtant hautement évitable avec une prise en charge particulière et multidisciplinaire. Pour qu'une prise en charge préventive soit efficace, il faudrait réussir à la détecter au minimum un mois avant.

### Méthodes médicales actuelles

Actuellement, les méthodes de détection de risque de survenue d'escarres utilisées en EHPAD sont des échelles d'évaluation qui ont été élaborées pour identifier les malades à haut risque et mettre en place une prévention. Ce sont des évaluations plus ou moins précises.

La plus connue est l'échelle de Norton (1962), qui a été validée pour des personnes de plus de 65 ans. Elle est simple

d'utilisation mais ne prend pas en compte le statut nutritionnel. On utilise aussi l'échelle de Braden (1985) qui est simple et claire (figure 2). Mais la simplicité de ces échelles est aussi un inconvénient, elles oublient des facteurs de risque importants (Kaewprag et al. 2015).

## Évaluation des résultats

Pour évaluer le pouvoir prédictif du modèle, le F-score est calculé. Il s'agit de la moyenne pondérée des scores de précision et de sensibilité (Chinchor 1992). La précision correspond au nombre de résidents dont on a correctement prédit la survenue d'escarre (les "vrais positifs") sur le nombre de résidents qui ont été prédit comme ayant des escarres (la somme des "vrais positifs" et des "faux positifs"). La sensibilité correspond aussi au nombre de résidents dont on a correctement prédit la survenue d'escarres mais sur le nombre de résidents qui ont réellement eu des escarres (la somme des "vrais positifs" et des "faux négatifs").

Le F-score a donc été obtenu avec la formule :

$$Fscore = 2 \times \frac{precision \times sensibilité}{precision + sensibilité}$$

$$= 2 \times \frac{\frac{vp}{vp+fp} \times \frac{vp}{vp+fn}}{\frac{vp}{vp+fp} + \frac{vp}{vp+fn}}$$

Nous essayons de maximiser le F-Score, soit de choisir le meilleur compromis entre Précision et Sensibilité. Pour de la prévention, il peut être plus important d'être plus sensible que précis, pour qu'un maximum de patients à risque soit pris en charge. Le choix d'un seuil qui prendrait en compte ce besoin est donc en discussion. Cependant, le F-score nous permet de comparer notre modèle avec d'autres méthodes de *Machine Learning* classique, mais aussi avec les méthodes de prévention actuellement utilisées en EHPAD (i.e. Norton et Braden)

Par ailleurs, l'utilisation de modèle graphique nous permet un autre type d'évaluation, contrairement aux autres méthodes. Ici, nous pouvons confronter l'explicabilité du modèle en étudiant le réseau bayésien obtenu et interroger un expert qui doit pouvoir y retrouver un sens médical.

Pour utiliser un classifieur par réseau bayésien, on apprend la structure du réseau à partir d'estimations sur la base d'apprentissage, ce qui fournit un outil efficace de calcul de la probabilité d'appartenir à une classe sachant les observations de chaque variable sur un individu. Or, une sélection de variables est possible avec cette méthode : si on connaît la couverture de Markov du réseau, elle suffit à effectuer la classification.

En effet, une couverture de Markov d'une variable aléatoire  $Y$  dans un ensemble de variables aléatoires  $\mathcal{S} = \{X_1, \dots, X_n\}$  se définit comme un sous-ensemble quelconque  $\mathcal{S}_1$  de  $\mathcal{S}$ , à condition que les autres variables soient indépendantes de  $Y$  :

$$Y \perp\!\!\!\perp \mathcal{S} \setminus \mathcal{S}_1 \mid \mathcal{S}_1.$$

Cela signifie que  $\mathcal{S}_1$  contient toutes les informations nécessaires pour l'inférence de  $Y$ , alors que les variables de  $\mathcal{S} \setminus \mathcal{S}_1$  sont redondantes. Plus particulièrement dans un réseau bayésien, la couverture de Markov d'un noeud précis contient ses parents, ses enfants et tous les autres parents de tous ses enfants (Pearl 1988).

## Automatisation de la classification

### Prétraitement

Pour ce projet, nous utilisons des données provenant de NETSoins, un logiciel complet sur l'accompagnement des résidents en EHPAD, édité par Teranga Software et utilisé par plus de 3 000 établissements, ce qui correspond à plus de 565 416 résidents dont 245 772 actuellement présents. Un nombre conséquent de données est donc potentiellement utilisable. NETSoins est accessible depuis un simple navigateur web. Il respecte le règlement général sur la protection des données (RGPD) et toutes les données sont centralisées et sécurisées dans des environnements d'Hébergement de Données de Santé certifié.

Les données sont extraites de façon complètement anonyme et proviennent de certains établissements non identifiables de toute la France. Toutes les dates sont exprimées en "délai après entrée" sans que l'on sache en quelle année le résident a séjourné dans l'EPHAD. Trois jeux de données distincts avec différentes temporalités pour les cas positifs sont créés : on regarde les données disponibles 1 mois avant le premier épisode d'escarre, 2 mois avant et 3 mois avant. Le jeu de données principal (1 mois avant) contient 116 702 résidents avec 11 934 porteurs d'escarres.

Il faut noter que le logiciel est utilisé et rempli par le personnel de l'établissement : les aides soignantes, infirmiers, médecins mais aussi les animateurs, les pharmaciens, le personnel administratif et paramédical ... La saisie des données concernant un résident en EHPAD est donc faite par des personnes distinctes à des moments différents plus ou moins périodiques. L'exploitation de ces données est donc difficile et ce travail est encore en cours. Pour l'instant, plusieurs techniques différentes sont utilisées, selon le type de données. Par exemple sur le GIR, un niveau de 1 à 6 qui indique la perte d'autonomie d'une personne âgée et qui n'est pas réévalué fréquemment, on crée 2 variables : pour le niveau à l'entrée, et le niveau actuel. Pour les événements ponctuels comme les chutes, on compte le nombre de chutes depuis l'entrée du résident dans l'établissement. Pour les pathologies, on vérifie la présence ou l'absence d'une pathologie au dernier moment disponible. Pour les données qui sont généralement actualisées de façon plus régulière, comme le poids, on les extrait avec le moment où le relevé de poids a été effectué et on calcule des tendances, ou des pourcentages d'évolution sur plusieurs durées déterminés. Ces méthodes ont été formulées en accord avec un professeur en Gériatrie.

Pour compléter les données manquantes, on utilise la

méthode `KNNImputer` de `scikit-learn`. Cela permet de compléter les valeurs manquantes en utilisant la méthode des 'k plus proches voisins'. On impute chacune des valeurs manquantes de l'échantillon en utilisant la valeur moyenne du k nombre de voisins le plus proche dans le jeu de données. Deux échantillons sont proches si leurs variables non manquantes sont proches. On utilise ici les 5 plus proches voisins (Troyanskaya et al. 2001).

Enfin, les variables continues ont été discrétisées en 10 catégories selon leurs déciles quand cela était nécessaire. Une discrétisation plus en accord avec les pratiques médicales est actuellement en cours.

### Apprentissage

Pour effectuer l'apprentissage de nos jeux de données, nous exploitons 75% de l'échantillon. Les 25% restants servent de base de test. La librairie `pyAgrum`, qui permet de construire des modèles et algorithmes pour les modèles graphiques probabilistes en Python (Ducamp, Gonzales, and Willemin 2020), a été utilisée. La fonction `BNClassifier` effectue un apprentissage automatique, sous la forme d'un classifieur probabiliste compatibles avec `scikit-learn`. Bien que le classifieur crée un réseau bayésien entier, il utilise de façon optimale uniquement les variables de la couverture de Markov.

De nombreux paramètres sont disponibles donc beaucoup de tests ont été effectués. Pour choisir le seuil de probabilité qui permet de savoir à quelle classe notre individu va être affecté, deux choix sont possibles. Tout d'abord, celui de la courbe (*Receiver Operating Characteristic*). Il s'agit du Taux de Vrais Positifs en fonction du Taux de Faux Positifs. Le seuil est donc le point pour lequel les deux axes sont maximisés. Le deuxième seuil est celui de la courbe PR (*Precision-Recall*), qui montre la Précision en fonction du Rappel (ou Sensibilité) et qui fonctionne sur le même principe. Nous avons choisi d'utiliser le seuil de la courbe PR car il prend en compte la prévalence des classes et est donc plus adapté à un échantillon avec des classes de tailles différentes. En effet, nous avons environ 10% de cas d'escarres pour 90% de cas témoins.

Nous pouvons aussi jouer sur la méthode d'apprentissage, comme l'algorithme `MIIC` (*Multivariate Information based Inductive Causation*) basé sur des contraintes (Affeldt and Isambert 2015; Verny et al. 2017) ou des méthodes plus classiques comme `Naive Bayes` ou `Greedy Hill Climbing` (Heckerman, Geiger, and Chickering 1995). La méthode d'apprentissage finalement utilisée est la méthode `Tabu` (Glover and Laguna 1998) car elle confère les meilleurs résultats avec nos données. Nous choisissons aussi un type de score, `BDeu` (Bayesian Dirichlet equivalent uniform) qui permet d'améliorer la qualité du réseau. Différents à priori ont aussi été testés sans que cela augmente nos résultats.

## Résultats

Les résultats obtenus sont disponibles dans la figure 3. On remarque pour notre `BNClassifier`, que plus la tem-

F-Score	1 mois	2 mois	3 mois
BN Classifier	0,70	0,69	0,67
Random Forest	0,72	0,69	0,70
AdaBoost	0,69	0,67	0,69
Nearest Neighbors	0,55	0,55	0,56
Logistic Regression	0,32	0,36	0,42
Braden	0,32	-	-
Norton	0,29	-	-

Figure 3: Tableau récapitulatif des différents F-Score selon les méthodes et les temporalités de prédiction.

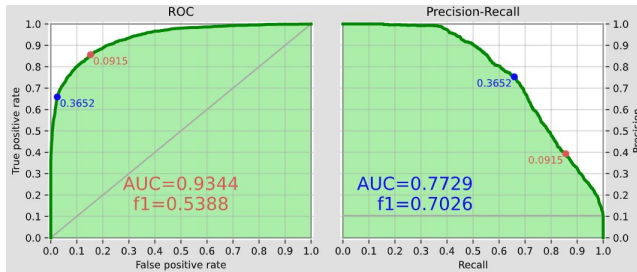


Figure 4: Courbes ROC et PR de notre BNClassifier avec le jeu de données 1 mois avant.

poralité est éloigné, plus il est difficile de prédire correctement l'escarre. Cependant, les résultats ne varient pas énormément et il faut prendre un compte le bénéfice sur le résident d'avoir une prise en charge préventive le plus tôt possible.

Une comparaison avec d'autres classifieurs de scikit-learn a été effectuée en figure 3. On observe qu'ils ne font généralement pas mieux que BNClassifier, sauf pour Random Forest, qui a un F-Score légèrement supérieur. Cependant, cette méthode ne nous apporte aucune information sur le modèle. On peut donc observer que la perte en terme de pouvoir prédictif n'est pas si conséquente. Quant à un modèle tel que la Logistic Regression, connue pour être explicable, ses résultats sont significativement plus bas que notre BNClassifier.

Par ailleurs, une rapide étude supplémentaire dans les données de NETSoins montre que les échelles de Norton et Braden sont peu utilisées (seulement 20% des résidents). De plus, leur F-Score, que nous avons calculés grâce aux données, sont effectivement mauvais : 0,32 pour l'échelle de Norton et 0,29 pour l'échelle de Braden. Il y a aussi un nombre important de faux négatifs, ce qui est le plus dangereux pour les résidents. Notre méthode utilisant les réseaux bayésiens a donc de bien meilleurs résultats que les méthodes actuellement utilisés en EHPAD et serait en plus calculé automatiquement, ce qui est un gain de temps conséquent pour le soignant.

Les courbe ROC et PR ont été tracées sur la figure 4.

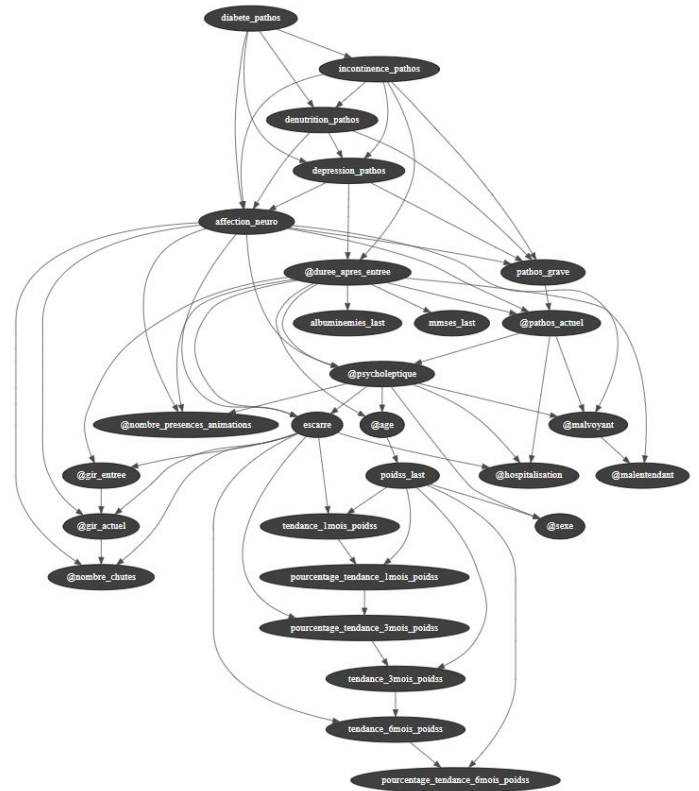


Figure 5: Le réseau bayésien de détection d'escarres obtenu.

Ainsi, en utilisant le seuil optimal de la courbe ROC (en rouge), le F-Score diminue grandement, bien que l'AUC (Area Under the Curve) soit très correcte. Le choix du seuil optimal de la courbe PR est donc pertinent.

Le réseau bayésien obtenu figure 5 a par ailleurs pu être confronté aux connaissances d'expert (professeur en Gériatrie). La couverture de Markov, en figure 6, a en particulier été étudiée et validée partiellement par les connaissances cliniques, c'est-à-dire que les explications fournies par le modèle ont satisfait l'expert. En effet, on observe sur cette grande couverture de Markov, des liens avec des facteurs de risques connus : comme le manque de mobilité qui s'exprime par le GIR, le poids des pathologies avec les affections neuro-psychiatriques et les hospitalisations ou encore la dénutrition visible avec les évolutions de poids.

## Conclusion

Dans cet article, nous proposons une classification à partir de réseaux bayésiens dans un contexte prédictif médical qui se révèle efficace au niveau des résultats et pertinente au niveau de l'explicabilité du modèle. De nombreuses améliorations sont encore possibles, notamment dans l'intégration de nouvelles variables et dans une meilleure exploitation de nos séries temporelles.

Une analyse des SHAP Values (*SHapley Additive exPlanations*), qui répondent à un problème classique de *Machine Learning* et permettrait d'augmenter encore l'explicabilité

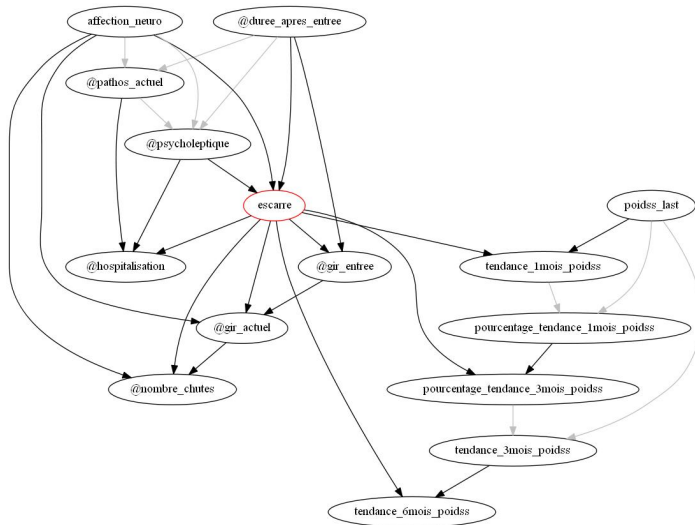


Figure 6: La couverture de Markov de la figure 5. En rouge, la variable centrale de la couverture de Markov et en gris, les arcs présents dans le réseau bayésien inutilisés dans la construction de la couverture de Markov.

de notre modèle, va aussi être effectué.

Dans un second temps, la prédiction devra être confrontée aux possibilités d'interventions. Il s'agira donc d'analyser la transformation d'un modèle prédictif à un modèle causal permettant le calcul de l'effet d'interventions. Une attention particulière sera apportée aux modèles permettant une explication des interventions recommandées (Pearl 2009, 2019; Bareinboim and Pearl 2016).

L'objectif est que l'approche soit mise en fonctionnement pour générer des alertes visant à attirer l'attention des médecins et des soignants sur des situations à haut risque et de les encourager à mener des interventions préventives des événements ciblés. Un essai clinique en cluster pourra être envisagé pour comparer les incidences des événements ciblés dans des établissements munis d'un tel système à des établissements contrôle n'en ayant pas. Enfin, l'approche pourra aussi être développée pour d'autres événements défavorables de santé, comme l'hospitalisation en urgence et la fracture du col du fémur.

## Remerciements

Ce travail a été effectué dans le cadre d'une thèse CIFRE (n°2019/1519) soutenue par Teranga Software, l'entreprise qui édite le logiciel NETSoins, et l'ANRT (Association Nationale de la Recherche et de la Technologie).

## References

Affeldt, S.; and Isambert, H. 2015. Robust Reconstruction of Causal Graphical Models based on Conditional 2-point and 3-point Information. In *ACI@ UAI*, 1–29.

Bareinboim, E.; and Pearl, J. 2016. Causal inference and

the data-fusion problem. *Proceedings of the National Academy of Sciences* 113(27): 7345–7352. ISSN 0027-8424, 1091-6490. doi:10.1073/pnas.1510507113. URL <https://www.pnas.org/content/113/27/7345>. Publisher: National Academy of Sciences Section: Colloquium Paper.

Belmin, J.; Chassagne, P.; Friocourt, P.; Gonthier, R.; Jeandel, C.; Nourhashemi, F.; and Pfitzenmeyer, P. 2016. *Gériatrie: pour le Praticien*. Elsevier Health Sciences. ISBN 978-2-294-75319-0.

Chinchor, N. 1992. MUC-4 Evaluation Metrics. In *Proceedings of the 4th Conference on Message Understanding*, MUC4 '92, 22–29. USA: Association for Computational Linguistics. ISBN 1558602739. doi:10.3115/1072064.1072067. URL <https://doi.org/10.3115/1072064.1072067>.

Dam, H. K.; Tran, T.; and Ghose, A. 2018. Explainable Software Analytics.

Ducamp, G.; Gonzales, C.; and Wuillemin, P.-H. 2020. aGrUM/pyAgrum : a toolbox to build models and algorithms for Probabilistic Graphical Models in Python. In *10th International Conference on Probabilistic Graphical Models*, volume 138 of *Proceedings of Machine Learning Research*, 609–612. Skørping, Denmark. URL <https://hal.archives-ouvertes.fr/hal-03135721>.

Glover, F.; and Laguna, M. 1998. *Tabu Search*, 2093–2229. Boston, MA: Springer US. ISBN 978-1-4613-0303-9. doi:10.1007/978-1-4613-0303-9.

HAS. 2001. Haute Autorité de Santé - Prévention et traitement des escarres de l'adulte et du sujet âgé.

Heckerman, D.; Geiger, D.; and Chickering, D. 1995. Learning Bayesian Networks: The Combination of Knowledge and Statistical Data. *Machine Learning* 20: 197–243. doi:10.1007/BF00994016.

Kaewprag, P.; Newton, C.; Vermillion, B.; Hyun, S.; Huang, K.; and Machiraju, R. 2015. Predictive Modeling for Pressure Ulcers from Intensive Care Unit Electronic Health Records. *AMIA Summits on Translational Science Proceedings* 2015: 82–86. ISSN 2153-4063. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4525237/>.

Koller, D.; and Friedman, N. 2009. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press. ISBN 978-0-262-01319-2. Google-Books-ID: 7dzpHCHzNQ4C.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. ISBN 978-1-55860-479-7.

Pearl, J. 2009. *Causality: Models, Reasoning and Inference (2nd ed.)*. Cambridge University Press. ISBN 978-1-139-64398-6. Google-Books-ID: LLkhAwwAAQBAJ.

Pearl, J. 2019. The seven tools of causal inference, with reflections on machine learning. *Communications of the ACM* 62(3): 54–60. ISSN 0001-0782. doi:10.1145/3241036. URL <https://doi.org/10.1145/3241036>.

Troyanskaya, O.; Cantor, M.; Sherlock, G.; Hastie, T.; Tibshirani, R.; Botstein, D.; and Altman, R. 2001. Missing

Value Estimation Methods for DNA Microarrays. *Bioinformatics* 17: 520–525. doi:10.1093/bioinformatics/17.6.520.

Verny, L.; Sella, N.; Affeldt, S.; Singh, P. P.; and Isambert, H. 2017. Learning causal networks with latent variables from multivariate information in genomic data. *PLoS computational biology* 13(10): e1005662.