



PRESENTATION FINALE BIBLIOGRAPHIE ET ETUDE DES CAS (B.E.C)

PREDICTIVE MODELING HOSPITAL READMISSION RATES USING
ELECTRONIC MEDICAL RECORD-WIDE MACHINE LEARNING: A CASE
STUDY USING MOUNT SINAI HEART FAILURE COHORT.
K SHAMEER, K.W.JOHNSON, A.YAHI, R.MIOTTO et. al.

MEDOU Daniel Magloire

Étudiant en fin de cycle Master recherche Informatique: Systèmes
Intelligents et Multimédia

Superviseur Académique : **Dr. HO TUONG VINH**

Année académique 2017 - 2018



PLAN

Introduction / Contexte

Résumé de l'article initial

État de l'art des articles connexes

Analyse

Conclusion et position personnelle

Références

CONTEXTE 1/3



Figure – 1. Retour des patients après hospitalisation initiale

CONTEXTE 2/3



Figure – 2. Patient réadmis à l'hospitalisation

CONTEXTE 3/3



Figure – 3. Étude pour limiter les cas de réadmission



PROBLÉMATIQUE

Face à ce genre de crise plusieurs questions sont posées



PROBLÉMATIQUE

Face à ce genre de crise plusieurs questions sont posées

- Quelle est la cause du retour des patients ?



PROBLÉMATIQUE

Face à ce genre de crise plusieurs questions sont posées

- Quelle est la cause du retour des patients ?
- Le traitement est - il approprié ?



PROBLÉMATIQUE

Face à ce genre de crise plusieurs questions sont posées

- Quelle est la cause du retour des patients ?
- Le traitement est - il approprié ?
- Les patients sont ils bien pris en charge ?



PROBLÉMATIQUE

Face à ce genre de crise plusieurs questions sont posées

- Quelle est la cause du retour des patients ?
- Le traitement est - il approprié ?
- Les patients sont ils bien pris en charge ?
- Comment procéder pour diminuer le taux de réadmission des patients atteints des affections chroniques ou aiguës aux USA ?



PROBLÉMATIQUE

Face à ce genre de crise plusieurs questions sont posées

- Quelle est la cause du retour des patients ?
- Le traitement est - il approprié ?
- Les patients sont ils bien pris en charge ?
- Comment procéder pour diminuer le taux de réadmission des patients atteints des affections chroniques ou aiguës aux USA ?

Modélisation Prédictive de Taux de Réadmission Hospitalière
au Moyen de l'Apprentissage de Machines Médicales à
Mémoire Électronique



RAPPEL ARTICLE INITIAL 1/3

Objectif principal

Mise en œuvre d'un modèle prédictif basé sur les données du DME pour prédire les taux de réadmission chez les patients souffrants d'insuffisance cardiaque.



RAPPEL ARTICLE INITIAL 1/3

Objectif principal

Mise en œuvre d'un modèle prédictif basé sur les données du DME pour prédire les taux de réadmission chez les patients souffrants d'insuffisance cardiaque.

Modalités utilisées

- Les codes de diagnostic.
- Les procédures.
- Les médicaments.
- Les mesures du laboratoire.
- Les signes vitaux.



RAPPEL ARTICLE INITIAL 1/3

Objectif principal

Mise en œuvre d'un modèle prédictif basé sur les données du DME pour prédire les taux de réadmission chez les patients souffrants d'insuffisance cardiaque.

Modalités utilisées

- Les codes de diagnostic.
- Les procédures.
- Les médicaments.
- Les mesures du laboratoire.
- Les signes vitaux.

RAPPEL ARTICLE INITIAL 2/3

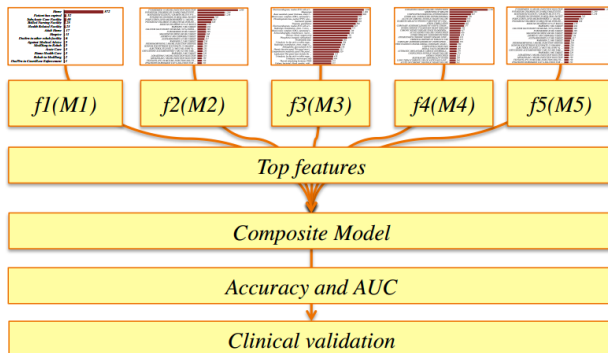


Figure – 4. Architecture d'apprentissage machine à l'échelle du DME et stratégie de modélisation prédictive pour trouver des moteurs de taux de réadmission

RAPPEL ARTICLE INITIAL 3/3

Au vue de ce tableau, Le modèle final est développé en utilisant 105 sur les 4205 caractéristiques avec une AUC = 0,78 et une précision de test de validation croisée de 83,19%.

Data-element	Type	Encoding	Accuracy	AUC	Features
Diagnosis	ICD-9 Diagnosis	Binary	70.3297%	0.605	34/1763
Procedures	ICD-9-Procedure	Binary	77.907%	<0.50	4/273
Procedure	CPT-codes	Binary	72.9858%	0.553	8/564
Medications	Medication name and dosage	Binary	81.9048%	0.615	26/1028
Labs	Non-descriptive lab measurements	Continuous	73.9336%	0.535	29/846
Composite model	Combined features	Hybrid	83.9000%	0.780	105

Figure – 5. Récapitulatif des différents prédicteurs et caractéristiques bayésiens compilés

Algorithme expérimenté : Naïve Bayes.



ÉTAT DE L'ART ARTICLE1 1/9

Titre : Impact des techniques de pré-traitement sélectionnées sur la prévision du risque de réadmission précoce chez les patients diabétiques en Inde.

Source : **International Journal of Diabetes in Developing Countries. Volume 36, October - December 2016. Pages 95 -102**



ÉTAT DE L'ART ARTICLE1 1/9

Titre : Impact des techniques de pré-traitement sélectionnées sur la prévision du risque de réadmission précoce chez les patients diabétiques en Inde.

Source : **International Journal of Diabetes in Developing Countries. Volume 36, October - December 2016. Pages 95 -102**

Objectif principal

Evaluer l'impacte des techniques de prétraitement sélectionnées sur la prédiction du risque de réadmission de 30 jours chez les patients diabétiques en Inde.



ÉTAT DE L'ART ARTICLE1 2/9

Critères appliqués pour extraire l'ensemble des données
nécessaire pour la visite des patients

- Admission à l'hôpital pour patients hospitalisés.



ÉTAT DE L'ART ARTICLE1 2/9

Critères appliqués pour extraire l'ensemble des données nécessaire pour la visite des patients

- Admission à l'hôpital pour patients hospitalisés.
- Durée du séjour (DDS) est minimum 1 jour. Garde de jour non incluse.



ÉTAT DE L'ART ARTICLE1 2/9

Critères appliqués pour extraire l'ensemble des données nécessaire pour la visite des patients

- Admission à l'hôpital pour patients hospitalisés.
- Durée du séjour (DDS) est minimum 1 jour. Garde de jour non incluse.
- Les patients de plus de 18 ans.



ÉTAT DE L'ART ARTICLE1 2/9

Critères appliqués pour extraire l'ensemble des données nécessaire pour la visite des patients

- Admission à l'hôpital pour patients hospitalisés.
- Durée du séjour (DDS) est minimum 1 jour. Garde de jour non incluse.
- Les patients de plus de 18 ans.
- Rencontres exclues en raison de l'accouchement.



ÉTAT DE L'ART ARTICLE1 2/9

Critères appliqués pour extraire l'ensemble des données nécessaire pour la visite des patients

- Admission à l'hôpital pour patients hospitalisés.
- Durée du séjour (DDS) est minimum 1 jour. Garde de jour non incluse.
- Les patients de plus de 18 ans.
- Rencontres exclues en raison de l'accouchement.
- Visite à l'hôpital d'un patient atteint de diabète.



ÉTAT DE L'ART ARTICLE1 3/9

Type et la nature des données

- 58 625 hospitalisations dans un hôpital indien.



ÉTAT DE L'ART ARTICLE1 3/9

Type et la nature des données

- 58 625 hospitalisations dans un hôpital indien.
- Diagnostic extrait, antécédents et procédure informations, médicaments et données de laboratoire. Rencontres du 1er juillet 2013 au 31 juillet 2015.



ÉTAT DE L'ART ARTICLE1 3/9

Type et la nature des données

- 58 625 hospitalisations dans un hôpital indien.
- Diagnostic extrait, antécédents et procédure informations, médicaments et données de laboratoire. Rencontres du 1er juillet 2013 au 31 juillet 2015.
- 18 616 admissions en garderie.



ÉTAT DE L'ART ARTICLE1 3/9

Type et la nature des données

- 58 625 hospitalisations dans un hôpital indien.
- Diagnostic extrait, antécédents et procédure informations, médicaments et données de laboratoire. Rencontres du 1er juillet 2013 au 31 juillet 2015.
- 18 616 admissions en garderie.
- 3 118 admission avec des patients de moins de 18 ans au moment de l'admission.



ÉTAT DE L'ART ARTICLE1 3/9

Type et la nature des données

- 58 625 hospitalisations dans un hôpital indien.
- Diagnostic extrait, antécédents et procédure informations, médicaments et données de laboratoire. Rencontres du 1er juillet 2013 au 31 juillet 2015.
- 18 616 admissions en garderie.
- 3 118 admission avec des patients de moins de 18 ans au moment de l'admission.
- 994 admissions avec diagnostic grossesse liée ou obstétrique



ÉTAT DE L'ART ARTICLE1 3/9

Type et la nature des données

- 58 625 hospitalisations dans un hôpital indien.
- Diagnostic extrait, antécédents et procédure informations, médicaments et données de laboratoire. Rencontres du 1er juillet 2013 au 31 juillet 2015.
- 18 616 admissions en garderie.
- 3 118 admission avec des patients de moins de 18 ans au moment de l'admission.
- 994 admissions avec diagnostic grossesse liée ou obstétrique



ÉTAT DE L'ART ARTICLE1 4/9

Type et la nature des données

- 26 516 patients n'ayant pas de diagnostic ou d'antécédents de DIABÈTE.



ÉTAT DE L'ART ARTICLE1 4/9

Type et la nature des données

- 26 516 patients n'ayant pas de diagnostic ou d'antécédents de DIABÈTE.
- **9 381 admissions qualifiées (patients diabétiques), 1 211 réadmis dans les 30 jours et 8 170 non réadmis dans les 30 jours**



ÉTAT DE L'ART ARTICLE1 5/9

Techniques de pré-traitement des données

- ① Sélection des caractéristiques.
 - L'approche par filtre basé sur la corrélation
 - Le test de Chi2 de Pearson



ÉTAT DE L'ART ARTICLE1 5/9

Techniques de pré-traitement des données

- ① Sélection des caractéristiques.
 - L'approche par filtre basé sur la corrélation
 - Le test de Chi2 de Pearson
- ② Imputation des valeurs manquantes.
 - La technique d'imputation moyenne pour les attributs numériques.
 - Le mode pour l'attribut nominal de tous les cas observés.



ÉTAT DE L'ART ARTICLE1 5/9

Techniques de pré-traitement des données

- ① Sélection des caractéristiques.
 - L'approche par filtre basé sur la corrélation
 - Le test de Chi2 de Pearson
- ② Imputation des valeurs manquantes.
 - La technique d'imputation moyenne pour les attributs numériques.
 - Le mode pour l'attribut nominal de tous les cas observés.
- ③ Réduction du déséquilibre des classes
 - La technique de sur-échantillonnage.
 - La technique de sous-échantillonnage.

ÉTAT DE L'ART ARTICLE1 6/9

Architecture

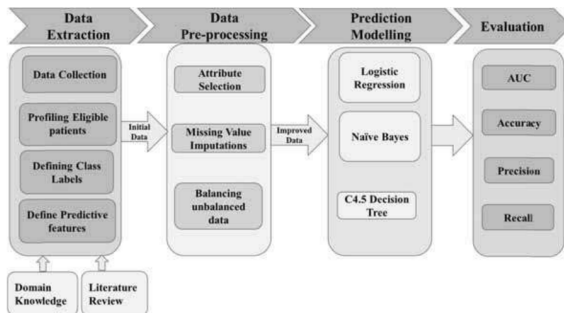


Figure – 6. L'architecture globale du processus de modélisation



ÉTAT DE L'ART ARTICLE1 7/9

Caractéristiques utilisées dans le processus de prédiction : 31 attributs

① Caractéristiques démographiques des patients

- Âge
- Sexe

② Autres caractéristiques

- Département
- Source d'admission
- Médicament pris contre le diabète
- La durée de séjour, etc...

③ Comorbidité

- Problème de coeur
- Hypertension
- Insuffisance rénale, etc...



ÉTAT DE L'ART ARTICLE1 8/9

Évaluation expérimentale

- 1 Outil utilisé : WEKA v3.6.12.



ÉTAT DE L'ART ARTICLE1 8/9

Évaluation expérimentale

- ① Outil utilisé : WEKA v3.6.12.
- ② Le test du modèle prédictif a été effectué en utilisant une procédure de validation croisée 10 fois supérieure.



ÉTAT DE L'ART ARTICLE1 8/9

Évaluation expérimentale

- ① Outil utilisé : WEKA v3.6.12.
- ② Le test du modèle prédictif a été effectué en utilisant une procédure de validation croisée 10 fois supérieure.
- ③ Quatre paramètres d'évaluation ont été considérés pour évaluer la qualité des différents modèles
 - Zone sous la courbe (AUC).
 - Precision.
 - Recall.
 - Accuracy.

ÉTAT DE L'ART ARTICLE1 9/9

Évaluation de la méthode de base et des techniques nouvellement proposées

Feature selection	Missing value imputation	Balancing	Predictive model	AUC	Precision	Recall	Accuracy
Baseline							
–	–	–	Logistic regression	0.68	0.56	0.08	0.87
–	–	–	Naïve Bayes	0.68	0.33	0.23	0.84
–	–	–	Decision tree	0.56	0.42	0.02	0.87
Top 3—AUC, precision, and recall							
Chi-square	Mean/mode imputation	Oversampling	Decision tree	0.86	0.82	0.85	0.83
Correlation-based	Mean/mode imputation	Oversampling	Decision tree	0.83	0.77	0.80	0.78
Chi-square	Mean/mode imputation	Oversampling	Logistic regression	0.83	0.74	0.78	0.75
Top 3—accuracy							
Chi-square	Mean/mode imputation	–	Logistic regression	0.68	0.56	0.08	0.87
Correlation-based	Mean/mode imputation	–	Naïve Bayes	0.66	0.56	0.10	0.87
Chi-square	Mean/mode imputation	–	Decision tree	0.58	0.48	0.04	0.87



ÉTAT DE L'ART ARTICLE2 1/6

Titre : Predicting 30-day Hospital Readmission with Publicly Available Administrative Database.

Source : **A Conditional Logistic Regression Modeling Approach. Big Data Analytics, Schattauer 2015**



ÉTAT DE L'ART ARTICLE2 1/6

Titre : Predicting 30-day Hospital Readmission with Publicly Available Administrative Database.

Source : **A Conditional Logistic Regression Modeling Approach. Big Data Analytics, Schattauer 2015**

Objectif principal

Explorer l'utilisation de la régression logistique conditionnelle pour augmenter la précision de la prédiction de réadmission hospitalière des patients atteints d'insuffisance cardiaque.



ÉTAT DE L'ART ARTICLE2 2/6

Méthodes

- ① Analyse des données sur les dossiers de congés pour patients hospitalisés. Les éléments retenus pour cette analyse.



ÉTAT DE L'ART ARTICLE2 2/6

Méthodes

- ① Analyse des données sur les dossiers de congés pour patients hospitalisés. Les éléments retenus pour cette analyse.
 - Les données démographiques des patients.



ÉTAT DE L'ART ARTICLE2 2/6

Méthodes

- ① Analyse des données sur les dossiers de congés pour patients hospitalisés. Les éléments retenus pour cette analyse.
 - Les données démographiques des patients.
 - Les données cliniques.



ÉTAT DE L'ART ARTICLE2 2/6

Méthodes

- ① Analyse des données sur les dossiers de congés pour patients hospitalisés. Les éléments retenus pour cette analyse.
 - Les données démographiques des patients.
 - Les données cliniques.
 - Les données sur l'utilisation des soins en Californie.



ÉTAT DE L'ART ARTICLE2 2/6

Méthodes

- ① Analyse des données sur les dossiers de congés pour patients hospitalisés. Les éléments retenus pour cette analyse.
 - Les données démographiques des patients.
 - Les données cliniques.
 - Les données sur l'utilisation des soins en Californie.
- ② Extraction des dossiers des bénéficiaires de l'assurance maladie ayant ne expérience hospitalière de 11 mois.



ÉTAT DE L'ART ARTICLE2 2/6

Méthodes

- ① Analyse des données sur les dossiers de congés pour patients hospitalisés. Les éléments retenus pour cette analyse.
 - Les données démographiques des patients.
 - Les données cliniques.
 - Les données sur l'utilisation des soins en Californie.
- ② Extraction des dossiers des bénéficiaires de l'assurance maladie ayant ne expérience hospitalière de 11 mois.
- ③ Correction du problème de déséquilibre des données avec un sous-échantillonnage.



ÉTAT DE L'ART ARTICLE2 3/6

Type et la nature des données

- Ensemble de 36 variables retenues pour le développement du modèle de prédiction.



ÉTAT DE L'ART ARTICLE2 3/6

Type et la nature des données

- Ensemble de 36 variables retenues pour le développement du modèle de prédiction.
- L'ensemble de données original contenait 3 970 921 enregistrements.



ÉTAT DE L'ART ARTICLE2 3/6

Type et la nature des données

- Ensemble de 36 variables retenues pour le développement du modèle de prédiction.
- L'ensemble de données original contenait 3 970 921 enregistrements.
- Après avoir effectué tous les éléments d'extraction de données, les dossiers de 22 410 patients sont restés.



ÉTAT DE L'ART ARTICLE2 3/6

Type et la nature des données

- Ensemble de 36 variables retenues pour le développement du modèle de prédiction.
- L'ensemble de données original contenait 3 970 921 enregistrements.
- Après avoir effectué tous les éléments d'extraction de données, les dossiers de 22 410 patients sont restés.
- 17 434 n'ont pas été réadmis dans les 30 jours suivant leur sortie. Classe majoritaire



ÉTAT DE L'ART ARTICLE2 3/6

Type et la nature des données

- Ensemble de 36 variables retenues pour le développement du modèle de prédiction.
- L'ensemble de données original contenait 3 970 921 enregistrements.
- Après avoir effectué tous les éléments d'extraction de données, les dossiers de 22 410 patients sont restés.
- 17 434 n'ont pas été réadmis dans les 30 jours suivant leur sortie. Classe majoritaire
- 4976 ont été réadmis dans les 30 jours suivant leur sortie.



ÉTAT DE L'ART ARTICLE2 4/6

Techniques appliquées

- ① Réduction du déséquilibre des classes
 - La technique de sur-échantillonnage.
 - La technique de sous-échantillonnage.



ÉTAT DE L'ART ARTICLE2 4/6

Techniques appliquées

- ① Réduction du déséquilibre des classes
 - La technique de sur-échantillonnage.
 - La technique de sous-échantillonnage.
- ② Le régression logistique standard.



ÉTAT DE L'ART ARTICLE2 4/6

Techniques appliquées

- ① Réduction du déséquilibre des classes
 - La technique de sur-échantillonnage.
 - La technique de sous-échantillonnage.
- ② Le régression logistique standard.
- ③ L'arbre de décisions. Avec comme objectifs



ÉTAT DE L'ART ARTICLE2 4/6

Techniques appliquées

- ① Réduction du déséquilibre des classes
 - La technique de sur-échantillonnage.
 - La technique de sous-échantillonnage.
- ② Le régression logistique standard.
- ③ L'arbre de décisions. Avec comme objectifs
 - Obtenir des variables influentes.



ÉTAT DE L'ART ARTICLE2 4/6

Techniques appliquées

- ① Réduction du déséquilibre des classes
 - La technique de sur-échantillonnage.
 - La technique de sous-échantillonnage.
- ② Le régression logistique standard.
- ③ L'arbre de décisions. Avec comme objectifs
 - Obtenir des variables influentes.
 - Déduire des règles de décisions.



ÉTAT DE L'ART ARTICLE2 4/6

Techniques appliquées

- ① Réduction du déséquilibre des classes
 - La technique de sur-échantillonnage.
 - La technique de sous-échantillonnage.
- ② Le régression logistique standard.
- ③ L'arbre de décisions. Avec comme objectifs
 - Obtenir des variables influentes.
 - Déduire des règles de décisions.
- ④ Ensemble de données stratifié et appliqué une régression logistique sur chaque strate de données.



ÉTAT DE L'ART ARTICLE2 4/6

Techniques appliquées

- ① Réduction du déséquilibre des classes
 - La technique de sur-échantillonnage.
 - La technique de sous-échantillonnage.
- ② Le régression logistique standard.
- ③ L'arbre de décisions. Avec comme objectifs
 - Obtenir des variables influentes.
 - Déduire des règles de décisions.
- ④ Ensemble de données stratifié et appliqué une régression logistique sur chaque strate de données.
- ⑤ Une validation croisée avait été appliquée dans le but d'évaluer la performance de prédiction globale de la régression Logistique Conditionnelle (CLR).



ÉTAT DE L'ART ARTICLE2 4/6

Techniques appliquées

- ① Réduction du déséquilibre des classes
 - La technique de sur-échantillonnage.
 - La technique de sous-échantillonnage.
- ② Le régression logistique standard.
- ③ L'arbre de décisions. Avec comme objectifs
 - Obtenir des variables influentes.
 - Déduire des règles de décisions.
- ④ Ensemble de données stratifié et appliqué une régression logistique sur chaque strate de données.
- ⑤ Une validation croisée avait été appliquée dans le but d'évaluer la performance de prédiction globale de la régression Logistique Conditionnelle (CLR).



ÉTAT DE L'ART ARTICLE2 5/6

Variables influentes identifiées par régression logistique

	Decision Rules	Total # of Subjects (# of admitted, # of not admitted)	Influential Variables identified for interactions from CLR2 ($p \leq 0.05$)	
			Variable Name	p-value
Stratum 1	DISPUniform = 5; NPR ≤ 4	1660 (1039, 621)	CM_COAG	0.04674
			CM_LYTE	0.02462
			DNR	0.00139
			FEMALE	0.01807
Stratum 2	DISPUniform = 5; NPR > 4	646 (462, 184)	RACE3 ^a	0.0455
Stratum 3	DISPUniform = 1 or 6; NCHRONIC ≤ 7	3278 (1309, 1969)	CM_LYTE	0.0245
			NPR	0.0173
			RACE2 ^b	0.0251
Stratum 4	DISUniform = 1 or 6; NCHRONIC > 7	4368 (2166, 2202)	CM_METS	0.00046
			CM_TUMOR	0.0207
			CM_WGTLOSS	0.0185
			DNR	0.0348
			MEDINCSTQ3 ^c	0.0458



ÉTAT DE L'ART ARTICLE2 6/6

Comparaison du modèle de classification

Classification Model		Prediction Accuracy*
LR: standard logistic regression		0.547
SLR: stepwise logistic regression		0.539
RF: random forests	all original variable	0.577
	only variables selected via SLR	0.574
SVM: support vector machine		0.560
CLR1: conditional logistic regression with 3 influence prediction variables	DISPUniform	0.548**
	NCHRONIC	0.564**
	NPR	0.576**
CLR2: conditional logistic regression with 4 data strata based on the first two layers of the decision tree		0.608**
CLR3: CLR2 + consideration of interacting variables based on identified influence ones in CLR2		0.615**



CONCLUSION

- Objectif est de faire une étude bibliographique d'un article
- Faire une étude des travaux connexes



CONCLUSION

- Objectif est de faire une étude bibliographique d'un article
- Faire une étude des travaux connexes
- Une étude comparative impossible car jeux de données n'étant pas formés des mêmes données ni soumis aux mêmes conditions. D'où résultats ont été individuellement présentés.



RÉFÉRENCES

Articles consultés

- ① K SHAMEER, K.W.JOHNSON, A.YAHI, R.MIOTTO et. al.
« Predictive Modeling of Hospital Readmission Rates Using Electronic Medical Record-wide Machine Learning : a case-study using Mount Sinai heart failure cohort ». Pacific Symposium on Biocomputing 2017. Page 1-10.



RÉFÉRENCES

Articles consultés

- ① K SHAMEER, K.W.JOHNSON, A.YAHI, R.MIOTTO et. al.
« Predictive Modeling of Hospital Readmission Rates Using Electronic Medical Record-wide Machine Learning : a case-study using Mount Sinai heart failure cohort ». Pacific Symposium on Biocomputing 2017. Page 1-10.
- ② R.Duggal, S.Shukla et al. International Journal of Diabetes in Developing Countries. Volume 36 Springer, October - December 2016. Pages 95 - 102.



RÉFÉRENCES

Articles consultés

- ① K SHAMEER, K.W.JOHNSON, A.YAHI, R.MIOTTO et. al.
« Predictive Modeling of Hospital Readmission Rates Using Electronic Medical Record-wide Machine Learning : a case-study using Mount Sinai heart failure cohort ». Pacific Symposium on Biocomputing 2017. Page 1-10.
- ② R.Duggal, S.Shukla et al. International Journal of Diabetes in Developing Countries. Volume 36 Springer, October - December 2016. Pages 95 - 102.
- ③ K.Zhu ; Z.Lou ; J.Zhou ; N.Ballester et al. Predicting 30-day Hospital Readmission with Publicly Available Administrative Database. A Conditional Logistic Regression Modeling Approach Schattauer 2015.

27 / 27